



IN THE UNITED STATES PATENT AND TRADEMARKS OFFICE

In re Application of:

Masato HORIE

Serial No.: 10/781,841

Art Unit: 1635

Filed: February 20, 2004

Examiner: WHITEMAN, BRIANA

For: HUMAN GENE

Declaration

Honorable Commissioner of Patents and Trademarks

Washington, D.C. 20231

SIR:

I, Masato HORIE declare that:

- 1) I am the inventor of the above-identified application, and am familiar with the subject matter of said application.
- 2) In order to demonstrate the utility of the present invention, the following experiments were carried out under my direction and supervision.

Experimental Data

Experiment 1. Test of the influence of recombinant NPR1 protein on cranial nerve growth activity

Since the mRNA and protein of NPR1 were observed to be highly expressed in the hippocampal region of the brain, this test examined the influence of recombinant rat NELL1 protein on survival of primary cultured neurons of rat hippocampus.

The reasons for using rat cells in this experiment are as follows:

- (1) use of human cells raises many ethical problems, but rat cells are free from such problems and are readily available;
- (2) although fresh primary cultured cells are most suitable as cells used for observing the development and extension of neurites, primary cultured human neurons cannot be easily prepared or obtained; and
- (3) rat NPR1 has a high sequence homology with human NPR1, and therefore

it can be presumed that the results obtained by using rat cells can be adequately extended to humans.

The DNA sequence of rat NPR1 mRNA is deposited in GenBank under accession number U48246, and its CDS (product="protein kinase C-binding protein NELL1") has about 93% homology with that of humans (Kang T et al., J. Bone and Mineral Research, Vol. 14, No. 1, pp. 80-89 (1999)).

1) Reagents etc. used in the test

- The anti-microtubule-associated protein-2 (MAP-2) mouse monoclonal antibody used as a neuron marker was purchased from Sternberger Monoclonals Inc.
- The anti-glial fibrillary acid protein (GFAP) antibody used as an astrocyte (nerve glial cells) marker was purchased from Sigma-Aldrich.
- The basic fibroblast growth factor (bFGF) was purchased from Upstate Biotechnology.
- The water-soluble tetrazolium salt was purchased from Dojindo Laboratories.
- Cell Counting Kit-8 for WST-8 assay, which contains WST-8 [2-(2-methoxy-4-nitrophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2H-tetrazolium, monosodium salt], was purchased from Dojindo Laboratories.

2) Production of recombinant rat NPR1protein (NELL1)

For production of the C-terminally FLAG-tagged NELL1 protein, a pIZT-mel-NELL-FLC plasmid was constructed by inserting the rat *NELL1* cDNA linked N-terminally to a mellitin signal peptide sequence and C-terminally to a FLAG epitope sequence into baculoviral vector pIZT/V5-His (Invitrogen).

High Five cells (BTI-TN-5B1-4, *Trichoplusia* nr.) were purchased from Invitrogen, and were cultured in High Five Serum-Free Medium (Invitrogen).

High Five cells were transfected with the pIZT-mel-Nell1-FLC plasmid using Cellfection (Invitrogen) according to the manufacturer's protocol. Forty-eight hours after transfection, cells were selected with 400 µg/mL of Zeocin (Invitrogen). The recombinant rat NELL1 protein was purified from the culture medium of Zeocin-resistant High Five cells by anion exchange chromatography

using a UNO Q-6 column (Bio-Rad).

3) Test method

3-1) The influence of NPR1 protein on survival of hippocampal cells (WST-8 assay)

Primary culturing of rat neurons was performed as follows, according to a method described in the literature (K. Abe, et al., "Effect of recombinant human basic fibroblast growth factor and its modified protein CS23 on survival of primary cultured neurons from various regions of fetal rat brain" Jpn. J. Pharmacol., 53, (1990), 221-227): the hippocampus was excised from the 18-day-old fetal Sprague-Dawley rat brain, and enzymatically digested with 0.25% trypsin and 0.002% DNaseI at 37°C for 15 minutes to obtain hippocampal cells. The obtained hippocampal cells were suspended in 10% FBS-containing DMEM medium, and inoculated into a poly-L-lysine-coated 96-well plate at a density of 3×10^5 cells/cm². On the next day, the medium was replaced with non-serum DMEM medium containing 1% N-2 supplements (Invitrogen), and culturing was performed for 3 days.

Neurons cannot survive in non-serum media, and gradually decrease in number as the number of days of culturing increases.

Predetermined amounts (1, 10, 100 or 1000 ng/mL) of NPR1 protein (the above purified NELL1) were added to the non-serum medium to continue culturing the hippocampal cells, and the survival of the cells after four-day culturing was determined by WST-8 assay.

3-2) The influence of NPR1 protein on survival of hippocampal cells (immunostaining)

Subsequently, the hippocampal cells after four-day culturing (cells cultured in the presence of 1000 ng/mL of NPR1 protein), and hippocampal cells cultured in a system without NELL1 for comparison (control) were fixed with 4% paraformaldehyde for immunohistological staining, permeabilized with 0.1% TritonX-100, blocked with 10% goat serum, and then washed with phosphate buffer.

Neurons and nerve glial cells (astrocytes) were identified with

anti-MAP2 antibody and anti-GFAP antibody, respectively, and counterstained with hematoxylin. The staining was visualized using Envision-labeled polymer reagent (DAKO) in combination with 3,3'-diaminobenzidine-tetrahydrochloride reagent.

4) Results

4-1) Results of WST-8 assay

Attached Fig. A is a graph showing survival (A450/650 nm) of the hippocampal cells after four-day culturing at each concentration of NELL1 used.

The results shown in Fig. A reveal that addition of NELL1 protein (purified NELL1) at a concentration of 10 ng/mL or more significantly enhances survival of hippocampal cells, compared to survival in the no-addition control ($p < 0.05$ vs control for 10ng/ml; $p < 0.01$ vs control for 100ng/ml and 1000ng/ml).

4-2) Results of immunostaining

Attached Fig. B is a set of stained images (photographs) of cells stained using anti-MAP-2 antibody, which is a neuron marker, and anti-GFAP antibody, which is an astrocyte marker. In Fig. B, the photograph on the upper left shows control hippocampal cells that were cultured in the absence of NELL1 and stained using anti-MAP-2 antibody (indicated as "Control (MAP-2)" in the figure); the photograph on the lower left shows control hippocampal cells that were cultured in the absence of NELL1 and stained using anti-GFAP antibody (indicated as "Control (GFAP)"); the photograph on the upper right shows hippocampal cells that were cultured in the presence of 1000 ng/mL of NELL1 and stained using anti-MAP-2 antibody (indicated as "NELL1 (MAP-2)"); and the photograph on the lower right shows hippocampal cells that were cultured in the presence of 1000 ng/mL of NELL1 and stained using anti-GFAP antibody (indicated as "NELL1 (GFAP)"). The bar (—) in each photograph is 100 μ m long.

Fig. B shows that, in the stained images obtained using anti-MAP-2 antibody, a greater number of stained cells and greater degree of development and extension of neurites are observed in the system containing NPR1 protein than in the control system containing no NPR1 protein.

In the stained images obtained using anti-GFAP antibody, there is no difference between the system containing NPR1 protein and the system containing no NPR1 protein.

5) Conclusions

The above experimental results demonstrate that NPR1 protein selectively enhances survival of neurons.

It was also demonstrated that NPR1 protein not only increases the survival of neurons, but also develops and extends the length of neurites.

As is evident from these results, NPR1 protein has nerve growth activity.

[Document Name]	Patent Application
[Reference Number]	2115JP
[Filing Date]	March 19, 1996
[Addressee]	The Commissioner of the Patent Office
[Int'l Classification]	C12A 15/11
[Title of the Invention]	HUMAN GENE
[Number of Claims]	3
[Inventor]	
[Address or Residence]	161-8, Aza Nakajima, Takashima, Naruto-cho, Naruto-shi, Tokushima- ken, Japan
[Name]	Tsutomu FUJIWARA
[Inventor]	
[Address or Residence]	Montemeiru Kawauchi 303, 79-1, Hiraishi Furuta, Kawauchi-cho, Tokushima-shi, Tokushima-ken, Japan
[Name]	Takeshi WATANABE
[Inventor]	
[Address or Residence]	Lavendar Haitsu 308, 4-44, Minamitamiya 4-chome, Tokushima- shi, Tokushima-ken, Japan
[Name]	Masato HORIE
[Inventor]	
[Address or Residence]	1-2-19-401, Nishisugamo, Toshima- ku, Tokyo-to, Japan
[Name]	Toyomasa KATAGIRI

[Patent Applicant]

[Identification Number] 000206956
[Address or Residence] 9, Kandatsukasacho 2-chome,
Chiyoda-ku, Tokyo-to, Japan
[Name] OTSUKA PHARMACEUTICAL CO., LTD.
[Representative] Akihiko OTSUKA

[Attorney]

[Identification Number] 100065215
[Patent Attorney]
[Name] Eiji SAEGUSA
[Telephone Number] 06-203-0941

[Attorney]

[Identification Number] 100076510
[Patent Attorney]
[Name] Hiromichi KAKEHI

[Attorney]

[Identification Number] 100086427
[Patent Attorney]
[Name] Takeshi OHARA

[Attorney]

[Identification Number] 100090066
[Patent Attorney]
[Name] Hiroshi NAKAGAWA

[Attorney]

[Identification Number] 100094101
[Patent Attorney]
[Name] Yasumitsu TACHI

[Indication of Fee]

[Deposit Account Number] 001616

[Fee Paid] 21000

[List of Document Filed]

[Document] Specification 1

[Document] Abstract 1

[Number of General Power of Attorney] 9101431

[Document Name] Specification

[Title of the Invention] HUMAN GENE

[CLAIMS]

5 [Claim 1] A GDP dissociation stimulating protein gene which comprises a nucleotide sequence coding for the amino acid sequence shown under SEQ ID NO:1.

[Claim 2] A GDP dissociation stimulating protein gene which comprises the nucleotide sequence shown under SEQ ID NO:2.

10 [Claim 3] A GDP dissociation stimulating protein gene as defined in Claim 2 which has the nucleotide sequence shown under SEQ ID NO:3.

[Detailed Description of the Invention]

[0001]

15 [Technical Field of the Invention]

The present invention relates to a gene useful as an indicator in the prophylaxis, diagnosis and treatment of diseases in humans. More particularly, it relates to a novel human gene analogous to rat, mouse, yeast, nematode and known human genes, among others, and
20 utilizable, after cDNA analysis thereof, chromosome mapping of cDNA and function analysis of cDNA, in gene diagnosis using said gene and in developing a novel therapeutic method.

25 [0002]

[Prior Art]

The genetic information of a living thing has been accumulated as sequences (DNA) of four bases, namely A, C, G and T, which exist in cell nuclei. Said genetic
5 information has been preserved for line preservation and ontogeny of each individual living thing.

[0003]

In the case of human being, the number of said bases is said to be about 3 billion (3×10^9) and
10 supposedly there are 50 to 100 thousand genes therein. Such genetic information serves to maintain biological phenomena in that regulatory proteins, structural proteins and enzymes are produced via such route that mRNA is transcribed from a gene (DNA) and then translated
15 into a protein. Abnormalities in said route from gene to protein translation are considered to be causative of abnormalities of life supporting systems, for example in cell proliferation and differentiation, hence causative of various diseases.

20 [0004]

As a result of gene analyses so far made, a number of genes which may be expected to serve as useful materials in drug development, have been found, for example genes for various receptors such as insulin
25 receptor and LDL receptor, genes involved in cell

proliferation and differentiation and genes for metabolic enzymes such as proteases, ATPase and superoxide dismutases.

[0005]

5 However, analysis of human genes and studies of the functions of the genes analyzed and of the relations between the genes analyzed and various diseases have been just begun and many points remain unknown. Further analysis of novel genes, analysis of the functions
10 thereof, studies of the relations between the genes analyzed and diseases, and studies for applying the genes analyzed to gene diagnosis or for medicinal purposes, for instance, are therefore desired in the relevant art.

[0006]

15 [Problems to be Solved by the Invention]

 If such a novel human gene as mentioned above can be provided, it will be possible to analyze the level of expression thereof in each cell and the structure and function thereof and, through expression product analysis
20 and other studies, it may become possible to reveal the pathogenesis of a disease associated therewith, for example a genopathy or cancer, or diagnose and treat said disease, for instance. It is an object of the present invention to provide such a novel human gene.

25 [0007]

For attaining the above object, the present inventors made intensive investigations and obtained the findings mentioned below. Based thereon, the present invention has now been completed.

5 [0008]

Thus, the present inventors synthesized cDNAs based on mRNAs extracted from various tissues, inclusive of human fetal brain, adult blood vessels and placenta, constructed libraries by inserting them into vectors, 10 allowing colonies of Escherichia coli transformed with said libraries to form on agar medium, picked up colonies at random and transferred to 96-well micro plates and registered a large number of human gene-containing E. coli clones.

15 [0009]

Each clone thus registered was cultivated on a small size, DNA was extracted and purified, the four base-specifically terminating extension reactions were carried out by the dideoxy chain terminator method using 20 the cDNA extracted as a template, and the base sequence of the gene was determined over about 400 bases from the 5' terminus thereof using an automatic DNA sequencer. Based on the thus-obtained base sequence information, a novel family gene analogous to known genes of animal and 25 plant species such as bacteria, yeasts, nematodes, mice

and humans was searched for.

[0010]

The method of the above-mentioned cDNA analysis is described in detail in the literature by Fujiwara, one
5 of the present inventors [Fujiwara Tsutomu, Saibo Kogaku (Cell Engineering), 14, 645-654 (1995)].

[0011]

Among this group, there are novel receptors, DNA binding domain-containing transcription regulating
10 factors, signal transmission system factors, metabolic enzymes and so forth. Based on the homology of the novel gene of the present invention as obtained by gene analysis to the genes analogous thereto, the product of the gene, hence the function of the protein, can
15 approximately be estimated by analogy. Furthermore, such functions as enzyme activity and binding ability can be investigated by inserting the candidate gene into an expression vector to give a recombinant.

[0012]

20 [Means for Solving the Problems]

According to the present invention, there are provided a novel human gene characterized by containing a nucleotide sequence coding for an amino acid sequence defined by SEQ ID NO:1, :4, :7, :10, :13, :16, :19, :22,
25 :25, :28, :31, :34, :37 or 40, a human gene characterized

by containing the nucleotide sequence defined by SEQ ID
NO:2, :5, :8, :11, :14, :17, :20, :23, :26, :29, :32,
:35, :38 or :41, respectively coding for the amino acid
sequence mentioned above, and a novel human gene

5 characterized by the nucleotide sequence defined by SEQ
ID NO:3, :6, :9, :12, :15, :18, :21, :24, :27, :30, :33,
:36, :39 or :42.

[0013]

The symbols used herein for indicating amino
10 acids, peptides, nucleotides, nucleotide sequences and so
on are those recommended by IUPAC and IUB or in "Guide-
line for drafting specifications, etc. including
nucleotide sequences or amino acid sequences" (edited by
the Japanese Patent Office), or those in conventional use
15 in the relevant field of art.

[0014]

As specific examples of such gene of the
present invention, there may be mentioned genes deducible
from the DNA sequences of the clones designated as "GEN-
20 501D08", "GEN-080G01", "GEN-025F07", "GEN-076C09", "GEN-
331G07", "GEN-163D09", "GEN-078D05TA13", "GEN-423A12",
"GEN-092E10", "GEN-428B12", "GEN-073E07", "GEN-093E05"
and "GEN-077A09" shown later herein in Examples 1 to 11.
The respective nucleotide sequences are as shown in the
25 sequence listing.

[0015]

These clones have an open reading frame comprising nucleotides (nucleic acid) respectively coding for the amino acids shown in the sequence listing. Their
5 molecular weights were calculated at the values shown later herein in the respective examples. Hereinafter, these human genes of the present invention are sometimes referred to as the designation used in Examples 1 to 11.

[0016]

10 [Mode of Carrying out the Invention]

In the following, the human gene of the present invention is described in further detail.

[0017]

As mentioned above, each human gene of the
15 present invention is analogous to rat, mouse, yeast, nematode and known human genes, among others, and can be utilized in human gene analysis based on the information about the genes analogous thereto and in studying the function of the gene analyzed and the relation between
20 the gene analyzed and a disease. It is possible to use said gene in gene diagnosis of the disease associated therewith and in exploitation studies of said gene for medicinal purposes.

[0018]

25 The gene of the present invention is

represented in terms of a single-stranded DNA sequence,
as shown under SEQ ID NO:2. It is to be noted, however,
that the present invention also includes a DNA sequence
complementary to such a single-stranded DNA sequence and
5 a component comprising both. The sequence of the gene of
the present invention as shown under SEQ ID NO: 3n-1
(where n is an integer of 1 to 14) is merely an example
of the codon combination encoding the respective amino
acid residues. The gene of the present invention is not
10 limited thereto but can of course have a DNA sequence in
which the codons are arbitrarily selected and combined
for the respective amino acid residues. The codon
selection can be made in the conventional manner, for
example taking into consideration the codon utilization
15 frequencies in the host to be used [Nucl. Acids Res., 9,
43-74 (1981)].

[0019]

The gene of the present invention further
includes DNA sequences coding for functional equivalents
20 derived from the amino acid sequence mentioned above by
partial amino acid or amino acid sequence substitution,
deletion or addition. These polypeptides may be produced
by spontaneous modification (mutation) or may be obtained
by posttranslational modification or by modifying the
25 natural gene (of the present invention) by a technique of

genetic engineering, for example by site-specific
mutagenesis [Methods in Enzymology, 154, p. 350, 367-382
(1987); ibid., 100, p. 468 (1983); Nucleic Acids
Research, 12, p. 9441 (1984); Zoku Seikagaku Jikken Koza
5 (Sequel to Experiments in Biochemistry) 1, "Idensi
Kenkyu-ho (Methods in Gene Research) II", edited by the
Japan Biochemical Society, p. 105 (1986)] or synthesizing
mutant DNAs by a chemical synthetic technique such as the
phosphotriester method or phosphoamidite method [J. Am.
10 Chem. Soc., 89, p. 4801 (1967); ibid., 91, p. 3350
(1969); Science, 150, p. 178 (1968); Tetrahedron Lett.,
22, p. 1859 (1981); ibid., 24, p. 245 (1983)], or by
utilizing the techniques mentioned above in combination.

[0020]

15 The protein encoded by the gene of the present
invention can be expressed readily and stably by
utilizing said gene, for example inserting it into a
vector for use with a microorganism and cultivating the
microorganism thus transformed.

20 [0021]

The protein obtained by utilizing the gene of
the present invention can be used in specific antibody
production. In this case, the protein producible in
large quantities by the genetic engineering technique
25 mentioned above can be used as the component to serve as

an antigen. The antibody obtained may be polyclonal or monoclonal and can be advantageously used in the purification, assay, discrimination or identification of the corresponding protein.

5 [0022]

The gene of the present invention can be readily produced based on the sequence information thereof disclosed herein by using general genetic engineering techniques [cf. e.g. Molecular Cloning, 2nd
10 Ed., Cold Spring Harbor Laboratory Press (1989); Zoku Seikagaku Jikken Koza, "Idenshi Kenkyu-ho I, II and III", edited by the Japan Biochemical Society (1986)].

[0023]

This can be achieved, for example, by selecting
15 a desired clone from a human cDNA library (prepared in the conventional manner from appropriate cells of origin in which the gene is expressed) using a probe or antibody specific to the gene of the present invention [e.g. Proc. Natl. Acad. Sci. USA, 78, 6613 (1981); Science, 222, 778
20 (1983)].

[0024]

The cells of origin to be used in the above method are, for example, cells or tissues in which the gene in question is expressed, or cultured cells derived
25 therefrom. Separation of total RNA, separation and

purification of mRNA, conversion to (synthesis of) cDNA, cloning thereof and so on can be carried out by conventional methods. cDNA libraries are also commercially available and such cDNA libraries, for example
5 various cDNA libraries available from Clontech Lab. Inc. can also be used in the above method.

[0025]

Screening of the gene of the present invention from these cDNA libraries can be carried out by the
10 conventional method mentioned above. These screening methods include, for example, the method comprising selecting a cDNA clone by immunological screening using an antibody specific to the protein produced by the corresponding cDNA, the technique of plaque or colony
15 hybridization using probes selectively binding to the desired DNA sequence, or a combination of these. As regards the probe to be used here, a DNA sequence chemically synthesized based on the information about the DNA sequence of the present invention is generally used.
20 It is of course possible to use the gene of the present invention or fragments thereof as the probe.

[0026]

Furthermore, a sense primer and an antisense primer designed based on the information about the
25 partial amino acid sequence of a natural extract isolated

and purified from cells or a tissue can be used as probes for screening.

[0027]

For obtaining the gene of the present
5 invention, the technique of DNA/RNA amplification by the
PCR method [Science, 230, 1350-1354 (1984)] can suitably
be employed. Particularly when the full-length cDNA can
hardly be obtained from the library, the RACE method
(rapid amplification of cDNA ends; Jikken Igaku
10 (Experimental Medicine), 12 (6), 35-38 (1994)], in
particular the 5'RACE method [Frohman, M. A., et al.,
Proc. Natl. Acad. Sci. USA, 85, 8998-9002 (1988)] is
preferably employed. The primers to be used in such PCR
method can be appropriately designed based on the
15 sequence information of the gene of the present invention
as disclosed herein and can be synthesized by a
conventional method.

[0028]

The amplified DNA/RNA fragment can be isolated
20 and purified by a conventional method as mentioned above,
for example by gel electrophoresis.

[0029]

The nucleotide sequence of the thus-obtained
gene of the present invention or any of various DNA
25 fragments can be determined by a conventional method, for

example the dideoxy method [Proc. Natl. Acad. Sci. USA, 74, 5463-5467 (1977)] or the Maxam-Gilbert method [Methods in Enzymology, 65, 499 (1980)]. Such nucleotide sequence determination can be readily performed using a
5 commercially available sequence kit as well.

[0030]

When the gene of the present invention is used and conventional techniques of recombinant DNA technology [see e.g. Science, 224, p. 1431 (1984); Biochem. Biophys. Res. Comm., 130, p. 692 (1985); Proc. Natl. Acad. Sci. USA, 80, p. 5990 (1983) and the references cited above]
10 are followed, a recombinant protein can be obtained. More detailedly, said protein can be produced by constructing a recombinant DNA enabling the gene of the present invention to be expressed in host cells,
15 introducing it into host cells for transformation thereof and cultivating the resulting transformant.

[0031]

In that case, the host cells may be eukaryotic or prokaryotic. The eukaryotic cells include vertebrate
20 cells, yeast cells and so on, and the vertebrate cells include, but are not limited to, simian cells named COS cells [Cell, 23, 175-182 (1981)], Chinese hamster ovary cells and a dihydrofolate reductase-deficient cell line
25 derived therefrom [Proc. Natl. Acad. Sci. USA, 77, 4216-

4220 (1980)] and the like, which are frequently used.

As regards the expression vector to be used with vertebrate cells, an expression vector having a promoter located upstream of the gene to be expressed, RNA splicing sites, a polyadenylation site and a transcription termination sequence can be generally used. This may further have an origin of replication as necessary. As an example of said expression vector, there may be mentioned pSV2dhfr [Mol. Cell. Biol., 1, 854 (1981)], which has the SV40 early promoter. As for the eukaryotic microorganisms, yeasts are generally and frequently used and, among them, yeasts of the genus Saccharomyces can be used with advantage. As regards the expression vector for use with said yeasts and other eukaryotic microorganisms, pAM82 [Proc. Natl. Acad. Sci. USA, 80, 1-5 (1983)], which has the acid phosphatase gene promoter, for instance, can be used.

[0032]

Furthermore, a prokaryotic gene fused vector can be preferably used as the expression vector for the gene of the present invention. As specific examples of said vector, there may be mentioned pGEX-2TK and pGEX-4T-2 which have a GST domain (derived from S. japonicum) with a molecular weight of 26,000.

[0033]

Escherichia coli and Bacillus subtilis are generally and preferably used as prokaryotic hosts. When these are used as hosts in the practice of the present invention, an expression plasmid derived from a plasmid vector capable of replicating in said host organisms and provided in this vector with a promoter and the SD (Shine and Dalgarno) sequence upstream of said gene for enabling the expression of the gene of the present invention and further provided with an initiation codon (e.g. ATG) necessary for the initiation of protein synthesis is preferably used. The Escherichia coli strain K12, among others, is preferably used as the host Escherichia coli, and pBR322 and modified vectors derived therefrom are generally and preferably used as the vector, while various known strains and vectors can also be used. Examples of the promoter which can be used are the tryptophan (trp) promoter, lpp promoter, lac promoter and PL/PR promoter.

[0034]

The thus-obtained desired recombinant DNA can be introduced into host cells for transformation by using various general methods. The transformant obtained can be cultured by a conventional method and the culture leads to expression and production of the desired protein encoded by the gene of the present invention. The medium

to be used in said culture can suitably be selected from among various media in conventional use according to the host cells employed. The host cells can be cultured under conditions suited for the growth thereof.

5 [0035]

In the above manner, the desired recombinant protein is expressed and produced and accumulated or secreted within the transformant cells or extracellularly or on the cell membrane.

10 [0036]

The recombinant protein can be separated and purified as desired by various separation procedures utilizing the physical, chemical and other properties thereof [cf. e.g. "Seikagaku (Biochemistry) Data Book II", pages 1175-1259, 1st Edition, 1st Printing, published June 23, 1980 by Tokyo Kagaku Dojin; Biochemistry, 25 (25), 8274-8277 (1986); Eur. J. Biochem., 163, 313-321 (1987)]. Specifically, said procedures include, among others, ordinary reconstitution treatment, treatment with a protein precipitating agent (salting out), centrifugation, osmotic shock treatment, sonication, ultrafiltration, various liquid chromatography techniques such as molecular sieve chromatography (gel filtration), adsorption chromatography, ion exchange chromatography, affinity chromatography and high-

15
20
25

performance liquid chromatography (HPLC), dialysis and combinations thereof. Among them, affinity chromatography utilizing a column with the desired protein bound thereto is particularly preferred.

5 [0037]

Furthermore, on the basis of the sequence information about the gene of the present invention as revealed by the present invention, for example by utilizing part or the whole of said gene, it is possible
10 to detect the expression of the gene of the present invention in various human tissues. This can be performed by a conventional method, for example by RNA amplification by RT-PCR (reverse transcribed-polymerase chain reaction) [Kawasaki, E. S., et al., Amplification
15 of RNA, in PCR Protocol, A guide to methods and applications, Academic Press, Inc., San Diego, 21-27 (1991)], or by northern blotting analysis [Molecular Cloning, Cold Spring Harbor Laboratory (1989)], with good results.

20 [0038]

The primers to be used in employing the above-mentioned PCR method are not limited to any particular ones provided that they are specific to the gene of the present invention and enable the gene of the present
25 invention alone to be specifically amplified. They can

be designed or selected appropriately based on the gene information provided by the present invention. They can have a partial sequence comprising about 20 to 30 nucleotides according to the established practice.

5 Suitable examples are as shown in Examples 1 to 11.

 [0039]

 Thus, the present invention also provides primers and/or probes useful in specifically detecting such novel gene.

10 [0040]

 [Effects of the Invention]

 By using the novel gene provided by the present invention, it is possible to detect the expression of said gene in various tissues, analyze the structure and
15 function thereof and, further, produce the human protein encoded by said gene in the manner of genetic engineering. These make it possible to analyze the expression product, reveal the pathology of a disease associated therewith, for example a genopathy or cancer,
20 and diagnose and treat the disease.

 [0041]

 [Examples]

 The following examples illustrate the present invention in further detail.

25 [0042]

[Example 1] GDP dissociation stimulator gene

(1) Cloning and DNA sequencing of GDP dissociation
stimulator gene

mRNAs extracted from the tissues of human fetal
5 brain, adult blood vessels and placenta were purchased
from Clontech and used as starting materials.

[0043]

cDNA was synthesized from each mRNA and
inserted into the vector λ ZAPII (Stratagene) to thereby
10 construct a cDNA library (Otsuka GEN Research Institute,
Otsuka Pharmaceutical Co., Ltd.)

[0044]

Human gene-containing Escherichia coli colonies
were allowed to form on agar medium by the in vivo
15 excision technique [Short, J. M., et al., Nucleic Acids
Res., 16, 7583-7600 (1988)]. Colonies were picked up at
random and human gene-containing Escherichia coli clones
were registered on 96-well micro plates. The clones
registered were stored at -80°C.

20 [0045]

Each of the clones registered was cultured
overnight in 1.5 ml of LB medium, and DNA was extracted
and purified using a model PI-100 automatic plasmid
extractor (Kurabo). Contaminant Escherichia coli RNA was
25 decomposed and removed by RNase treatment. The DNA was

dissolved to a final volume of 30 μ l. A 2- μ l portion was used for roughly checking the DNA size and quantity using a minigel, 7 μ l was used for sequencing reactions and the remaining portion (21 μ l) was stored as plasmid DNA at
5 4°C.

[0046]

This method, after slight changes in the program, enables extraction of the cosmid, which is useful also as a probe for FISH (fluorescence in situ
10 hybridization) shown later in the examples.

[0047]

Then, the dideoxy terminator method of Sanger et al. [Sanger, F., et al., Proc. Natl. Acad. Sci. USA, 74, 5463-5467 (1977)] using T3, T7 or a synthetic
15 oligonucleotide primer or the cycle sequence method [Carothers, A. M., et al., Bio. Techniques, 7, 494-499 (1989)] comprising the dideoxy chain terminator method plus PCR method was carried out. These are methods of terminating the extension reaction specifically to the
20 four bases using a small amount of plasmid DNA (about 0.1 to 0.5 μ g) as a template.

[0048]

The sequence primers used were FITC (fluorescein isothiocyanate)-labeled ones. Generally,
25 about 25 cycles of reaction were performed using Taq

polymerase. The PCR products were separated on a polyacrylamide urea gel and the fluorescence-labeled DNA fragments were submitted to an automatic DNA sequencer (ALFTM DNA Sequencer; Pharmacia) for determining the
5 sequence of about 400 bases from the 5' terminus side of cDNA.

[0049]

Since the 3' nontranslational region is high in heterogeneity for each gene and therefore suited for
10 discriminating individual genes from one another, sequencing was performed on the 3' side as well depending on the situation.

[0050]

The vast sum of nucleotide sequence information
15 obtained from the DNA sequencer was transferred to a 64-bit DEC 3400 computer for homology analysis by the computer. In the homology analysis, a data base (GenBank, EMBL) was used for searching according to the UWGCG FASTA program [Pearson, W. R. and Lipman, D. J.,
20 Proc. Natl. Acad. Sci. USA, 85, 2444-2448 (1988)].

[0051]

As a result of arbitrary selection by the above method and of cDNA sequence analysis, a clone designated as GEN-501D08 and having a 0.8 kilobase insert was found
25 to show a high level of homology to the C terminal region

of the human Ral guanine nucleotide dissociation
stimulator (RalGDS) gene. Since RalGDS is considered to
play a certain role in signal transmission pathways, the
whole nucleotide sequence of the cDNA insert portion
5 providing the human homolog was further determined.

[0052]

Low-molecular GTPases play an important role in
transmitting signals for a number of cell functions
including cell proliferation, differentiation and
10 transformation [Bourne, H. R. et al., Nature, 348, 125-
132 (1990); Bourne et al., Nature, 349, 117-127 (1991)].

[0053]

It is well known that, among them, those
proteins encoded by the ras gene family function as
15 molecular switches or, in other words, the functions of
the ras gene family are regulated by different conditions
of binding proteins such as biologically inactive GDP-
binding proteins or active GDP-binding proteins, and that
these two conditions are induced by GTPase activating
20 proteins (GAPs) or GDS. The former enzymes induce GDP
binding by stimulating the hydrolysis of bound GTP and
the latter enzyme induces the regular GTP binding by
releasing bound GDP [Bogusuki, M. S. and McCormick, F.,
Nature, 366, 643-654 (1993)].

25

[0054]

RalGDS was first discovered as a member of the ras gene family lacking in transforming activity and as a GDP dissociation stimulator specific to RAS [Chardin, P. and Tavittian, A., EMBO J., 5, 2203-2208 (1986); Albright, C. F., et al., EMBO J., 12, 339-347 (1993)].

[0055]

In addition to Ral, RalGDS was found to function, through interaction with these proteins, as an effector molecule for N-ras, H-ras, K-ras and Rap [Spaargaren, M. and Bischoff, J. R., Proc. Natl. Acad. Sci. USA, 91, 12609-12613 (1994)].

The nucleotide sequence of the cDNA clone designated as GEN-501D08 is shown under SEQ ID NO:3, the nucleotide sequence of the coding region of said clone under SEQ ID NO:2, and the amino acid sequence encoded by said nucleotide sequence under SEQ ID NO:1.

[0056]

This cDNA comprises 842 nucleotides, including an open reading frame comprising 366 nucleotides and coding for 122 amino acids. The translation initiation codon was found to be located at the 28th nucleotide residue.

[0057]

Comparison between the RalGDS protein known among conventional databases and the amino acid sequence

deduced from said cDNA revealed that the protein encoded
by this cDNA is homologous to the C terminal domain of
human RalGDS. The amino acid sequence encoded by this
novel gene was found to be 39.5% identical with the C
5 terminal domain of RalGDS which is thought to be
necessary for binding to ras.

[0058]

Therefore, it is presumable, as mentioned
above, that this gene product might interact with the ras
10 family proteins or have influence on the ras-mediated
signal transduction pathways. However, this novel gene
is lacking in the region coding for the GDS activity
domain and the corresponding protein seems to be
different in function from the GDS protein. This gene
15 was named human RalGDS by the present inventors.

[0059]

(2) Northern blot analysis

The expression of the RalGDS protein mRNA in
normal human tissues was evaluated by Northern blotting
20 using, as a probe, the human cDNA clone labeled by the
random oligonucleotide priming method.

[0060]

The Northern blot analysis was carried out with
a human MTN blot (Human Multiple Tissue Northern blot;
25 Clontech, Palo Alto, CA, USA) according to the manufac-

turer's protocol.

[0061]

Thus, the PCR amplification product from the
above GEN-501D08 clone was labeled with [³²P]-dCTP
5 (random-primed DNA labeling kit, Boehringer-Mannheim) for
use as a probe.

[0062]

For blotting, hybridization was performed
overnight at 42°C in a solution comprising 50%
10 formamide/5 x SSC/50 x Denhardt's solution/0.1% SDS
(containing 100 µg/ml denatured salmon sperm DNA). After
washing with two portions of 2 x SSC/0.01% SDS at room
temperature, the membrane filter was further washed three
times with 0.1 x SSC/0.05% SDS at 50°C for 40 minutes.
15 An X-ray film (Kodak) was exposed to the filter at -70°C
for 18 hours.

[0063]

As a result, it was revealed that a 900-bp
transcript had been expressed in all the human tissues
20 tested. In addition, a 3.2-kb transcript was observed
specifically in the heart and skeletal muscle. The
expression of these transcripts differing in size may be
due either to alternative splicing or to cross
hybridization with homologous genes.

25

[0064]

(3) Cosmid clone and chromosome localization by FISH

FISH was performed by screening a library of human chromosomes cloned in the cosmid vector pWE15 using, as a probe, the 0.8-kb insert of the cDNA clone
5 [Sambrook, J., et al., Molecular Cloning, 2nd Ed., pp. 3.1-3.58, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)].

[0065]

FISH for chromosome assignment was carried out
10 by the method of Inazawa et al. which comprises G-banding pattern comparison for confirmation [Inazawa, J., et al., Genomics, 17, 153-162 (1993)].

[0066]

For use as a probe, the cosmid DNA (0.5 μ g)
15 obtained from chromosome screening and corresponding to GEN-501D08 was labeled with biotin-16-dUTP by nick translation.

[0067]

To eliminate the background noise due to
20 repetitive sequences, 0.5 μ l of sonicated human placenta DNA (10 mg/ml) was added to 9.5 μ l of the probe solution. The mixture was denatured at 80°C for 5 minutes and admixed with an equal volume of 4 x SSC containing 20% dextran sulfate. Then, a denatured slide was sown with
25 the hybridization mixture and, after covering with

paraffin, incubated in a wet chamber at 37°C for 16 to 18 hours. After washing with 50% formamide/2 x SSC at 37°C for 15 minutes, the slide was washed with 2 x SSC for 15 minutes and further with 1 x SSC for 15 minutes.

5 [0068]

The slide was then incubated in 4 x SSC supplemented with "1% Block Ace" (trademark; Dainippon Pharmaceutical) containing avidin-FITC (5 µg/ml) at 37°C for 40 minutes. Then, the slide was washed with 4 x SSC for 10
10 minutes and with 4 x SSC containing 0.05% Triton X-100 for 10 minutes and immersed in an antifading PPD solution [prepared by adjusting 100 mg of PPD (Wako Catalog No. 164-015321) and 10 ml of PBS(-) (pH 7.4) to pH 8.0 with 0.5 M Na₂CO₃/0.5 M NaHCO₃ (9:1, v/v) buffer (pH 9.0) and
15 adding glycerol to make a total volume of 100 ml] containing 1% DABCO [1% DABCO (Sigma) in PBS(-):glycerol 1:9 (v:v)], followed by counter staining with DAPI (4,6-diamino-2-phenylindole; Sigma).

[0069]

20 With more than 100 tested cells in the metaphase, a specific hybridization signal was observed on the chromosome band at 6p21.3, without any signal on other chromosomes. It was thus confirmed that the RalGDS gene is located on the chromosome 6p21.3.

25 [0070]

By using the novel human RalGDS-associated gene of the present invention as obtained in this example, the expression of said gene in various tissues can be detected and the human RalGDS protein can be produced in the manner of genetic engineering. These are expected to enable studies on the roles of the expression product protein and ras-mediated signals in transduction pathways as well as pathological investigations of diseases in which these are involved, for example cancer, and the diagnosis and treatment of such diseases. Furthermore, it becomes possible to study the development and progress of diseases involving the same chromosomal translocation of the RalGDS protein gene of the present invention, for example tonic spondylitis, atrial septal defect, pigmentary retinopathy, aphasia and the like.

[0071]

[Example 2] Cytoskeleton-associated protein 2 gene
(CKAP2 gene)

(1) Cytoskeleton-associated protein 2 gene cloning and
DNA sequencing

cDNA clones were arbitrarily chosen from a human fetal brain cDNA library in the same manner as in Example 1-(1) were subjected to sequence analysis and, as a result, a clone having a base sequence containing the CAP-glycine domain of the human cytoskeleton-associated

protein (CAP) gene and highly homologous to several CAP family genes was found and named GEN-080G01.

[0072]

Meanwhile, the cytoskeleton occurs in the
5 cytoplasm and just inside the cell membrane of eukaryotic
cells and is a network structure comprising complicatedly
entangled filaments. Said cytoskeleton is constituted of
microtubules composed of tubulin, microfilaments composed
of actin, intermediate filaments composed of desmin and
10 vimentin, and so on. The cytoskeleton not only acts as
supportive cellular elements but also isokinetically
functions to induce morphological changes of cells by
polymerization and depolymerization in the fibrous
system. The cytoskeleton binds to intracellular
15 organelles, cell membrane receptors and ion channels and
thus plays an important role in intracellular movement
and locality maintenance thereof and, in addition, is
said to have functions in activity regulation and mutual
information transmission. Thus it supposedly occupies a
20 very important position in physiological activity
regulation of the whole cell. In particular, the
relation between canceration of cells and qualitative
changes of the cytoskeleton attracts attention since
cancer cells differ in morphology and recognition
25 response from normal cells.

[0073]

The activity of this cytoskeleton is modulated by a number of cytoskeleton-associated proteins (CAPs). One group of CAPs is characterized by a glycine motif highly conserved and supposedly contributing to association with microtubules [CAP-GLY domain; Riehemann, K. and Song, C., Trends Biochem. Sci., 18, 82-83 (1993)].

[0074]

Among the members of this group of CAPs, there are CLIP-170, 150 kDa DAP (dynein-associated protein, or dynactin), D. melanogaster GLUED, S. cerevisiae BIK1, restin [Bilbe, G., et al., EMBO J., 11, 2103-2113 (1992)]; Hilliker, C., et al., Cytogenet. Cell Genet., 65, 172-176 (1994)] and C. elegans 113.5 kDa protein [Wilson, R., et al., Nature, 368, 32-38 (1994)]. Except for the last two proteins, direct or indirect evidences have suggested that they could interact with microtubules.

[0075]

The above-mentioned CLIP-170 is essential for the in vitro binding of endocytic vesicles to microtubules and colocalizes with endocytic organelles [Rickard, J. E. and Kreis, T. E., J. Biol. Chem., 18, 82-83 (1990); Pierre, P., et al., Cell, 70, 887-900 (1992)].

[0076]

The above-mentioned dynactin is one of the factors constituting the cytoplasmic dynein motor, which functions in retrograde vesicle transport [Schroer, T. A. and Sheetz, M. P., J. Cell Biol., 115, 1309-1318 (1991)] or probably in the movement of chromosomes during mitosis [Pfarr, C. M., et al., Nature, 345, 263-265 (1990); Steuer, E. R., et al., Nature, 345, 266-268 (1990); Wordeman, L., et al., J. Cell Biol., 114, 285-294 (1991)].

10 [0077]

GLUED, the Drosophila homolog of mammalian dynactin, is essential for the viability of almost all cells and for the proper organization of some neurons [Swaroop, A., et al., Proc. Natl. Acad. Sci. USA, 84, 6501-6505 (1987); Holzbaur, E. L. P., et al., Nature, 351, 579-583 (1991)].

[0078]

BIK1 interacts with microtubules and plays an important role in spindle formation during mitosis in yeasts [Trueheart, J., et al., Mol. Cell. Biol., 7, 2316-2326 (1987); Berlin, V., et al., J. Cell Biol., 111, 2573-2586 (1990)].

[0079]

At present, these genes are classified under the term CAP family (CAPs).

[0080]

As a result of database searching, the above-mentioned cDNA clone of 463-bp (excluding the poly-A signal) showed significant homology in nucleotide
5 sequence with the restin and CLIP-170 encoding genes. However, said clone was lacking in the 5' region as compared with the restin gene and, therefore, the technique of 5' RACE [Frohman, M. A., et al., Proc. Natl. Acad. Sci. USA, 8, 8998-9002 (1988)] was used to isolate
10 this missing segment.

[0081]

(2) 5' RACE (5' rapid amplification of cDNA ends)

A cDNA clone containing the 5' portion of the gene of the present invention was isolated for analysis
15 by the 5' RACE technique using a commercial kit (5'-Rapid AmpliFinder RACE kit, Clontech) according to the manufacturer's protocol with minor modifications, as follows.

[0082]

20 The gene-specific primer P1 and primer P2 used here were synthesized by the conventional method and their nucleotide sequences are as shown below in Table 1. The anchor primer used was the one attached to the commercial kit.

25 [0083]

[Table 1]

	Primer	Nucleotide sequence
5	Primer P1	5'-ACACCAATCCAGTAGCCAGGCTTG-3'
	Primer P2	5'-CACTCGAGAATCTGTGAGACCTACATACATGACG-3'

[0084]

10 cDNA was obtained by reverse transcription of
0.1 μ g of human fetal brain poly(A)+RNA by the random
hexamer technique using reverse transcriptase
(SuperscriptTM II, Life Technologies) and the cDNA was
15 amplified by the first PCR using the P1 primer and anchor
primer according to Watanabe et al. [Watanabe, T., et
al., Cell Genet., in press).

[0085]

Thus, to 0.1 μ g of the above-mentioned cDNA
were added 2.5 mM dNTP/1 x Taq buffer (Takara Shuzo)/0.2
20 μ M P1 primer, 0.2 μ M adaptor primer/0.25 unit ExTaq
enzyme (Takara Shuzo) to make a total volume of 50 μ l,
followed by addition of the anchor primer. The mixture
was subjected to PCR. Thus, 35 cycles of amplification
were performed under the conditions: 94°C for 45 seconds,
25 60°C for 45 seconds, and 72°C for 2 minutes. Finally,
the mixture was heated at 72°C for 5 minutes.

[0086]

Then, 1 μ l of the 50- μ l first PCR product was

subjected to amplification by the second PCR using the specific nested P2 primer and anchor primer. The second PCR product was analyzed by 1.5% agarose gel electrophoresis.

5 [0087]

Upon agarose gel electrophoresis, a single band, about 650 nucleotides in size, was detected. The product from this band was inserted into a vector (pT7Blue(R)T-Vector, Novagen) and a plurality of clones with an insert having an appropriate size were selected.

[0088]

Six of the 5' RACE clones obtained from the PCR product had the same sequence but had different lengths. By sequencing two overlapping cDNA clones, GEN-080G01 and GEN-080G0149, the protein-encoding sequence and 5' and 3' flanking sequences, 1015 nucleotides in total length, were determined. Said gene was named cytoskeleton-associated protein 2 gene (CKAP2 gene).

[0089]

20 The nucleotide sequence obtained from the above-mentioned two overlapping cDNA clones GEN-080G01 and GEN-080G0149 is shown under SEQ ID NO:6, the nucleotide sequence of the coding region of said clone under SEQ ID NO:5, and the amino acid sequence encoded by said nucleotide sequence under SEQ ID NO:4.

[0090]

As shown under SEQ ID NO:6, the CKAP2 gene had a relatively GC-rich 5' noncoding region, with incomplete triplet repeats, (CAG)₄(CGG)₄(CTG)(CGG), occurring at
5 nucleotides 40-69.

[0091]

ATG located at nucleotides 274-276 is the presumable start codon. A stop codon (TGA) was situated at nucleotides 853-855. A polyadenylation signal
10 (ATTAAA) was followed by 16 nucleotides before the poly(A) start. The estimated open reading frame comprises 579 nucleotides coding for 193 amino acid residues with a calculated molecular weight of 21,800 daltons.

15 [0092]

The coding region was further amplified by RT-PCR, to eliminate the possibility of the synthetic sequence obtained being a cDNA chimera.

[0093]

20 (2) Similarity of CKAP2 to other CAPs

While sequencing of CKAP2 revealed homology with the sequences of restin and CLIP-170, the homologous region was limited to a short sequence corresponding to the CAP-GLY domain. On the amino acid level, the deduced
25 CKAP2 was highly homologous to five other CAPs in this

domain.

[0094]

CKAP2 was lacking in such other motif characteristics of some CAPs as the alpha helical rod and zinc finger motif. The alpha helical rod is thought to contribute to dimerization and to increase the microtubule binding capacity [Pierre, P., et al., Cell, 70, 887-900 (1992)]. The lack of the alpha helical domain might mean that CKAP2 be incapable of homo or hetero dimer formation.

[0095]

Paralleling of the CAP-GLY domains of these proteins revealed that other conserved residues other than glycine residues are also found in CKAP2. CAPs having a CAP-GLY domain are thought to be associated with the activities of cellular organelles and the interactions thereof with microtubules. Since it contains a CAP-GLY domain, as mentioned above, CKAP2 is placed in the family of CAPs.

[0096]

Studies with mutants of Glued have revealed that the Glued product plays an important role in almost all cells [Swaroop, A., et al., Proc. Natl. Acad. Sci. USA, 84, 6501-6505 (1987)] and that it has other neuron-specific functions in neuronal cells [Meyerowitz, E. M.

and Kankel, D. R., Dev. Biol., 62, 112-142 (1978)].

These microtubule-associated proteins are thought to function in vesicle transport and mitosis. Because of the importance of the vesicle transport system in neuronal cells, defects in these components might lead to aberrant neuronal systems.

[0097]

In view of the above, CKAP2 might be involved in specific neuronal functions as well as in fundamental cellular functions.

[0098]

(3) Northern blot analysis

The expression of human CKAP2 mRNA in normal human tissues was examined by Northern blotting in the same manner as in Example 1-(2) using the GEN-080G01 clone (corresponding to nucleotides 553-1015) as a probe.

[0099]

As a result, in all the eight tissues tested, namely human heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas, a 1.0 kb transcript agreeing in size with the CKAP2 cDNA was detected. Said 1.0 kb transcript was expressed at significantly higher levels in heart and brain than in the other tissues examined. Two weak bands, 3.4 kb and 4.6 kb, were also detected in all the tissues examined.

[0100]

According to the Northern blot analysis, the 3.4 kb and 4.6 kb transcripts might possibly be derived from the same gene coding for the 1.0 kb CKAP2 by
5 alternative splicing or transcribed from other related genes. These characteristics of the transcripts may indicate that CKAP2 might also code for a protein having a CAP-GLY domain as well as an alpha helix.

[0101]

10 (4) Cosmid cloning and chromosomal localization by direct R-banding FISH

Two cosmids corresponding to the CKAP2 cDNA were obtained. These two cosmid clones were subjected to direct R-banding FISH in the same manner as in Example 1-
15 (3) for chromosomal locus mapping of CKAP2.

[0102]

For suppressing the background due to repetitive sequences, a 20-fold excessive amount of human Cot-I DNA (BRL) was added as described by Lichter et al.
20 [Lichter, P., et al., Proc. Natl. Acad. Sci. USA, 87, 6634-6638 (1990)]. A Provia 100 film (Fuji ISO 100; Fuji Photo Film) was used for photomicrography.

[0103]

As a result, CKAP2 was mapped on chromosome
25 bands 19q13.11-q13.12.

[0104]

Two autosomal dominant neurological diseases have been localized to this region by linkage analysis: CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) between the
5 DNA markers D19S221 and D19S222, and FHM (familial hemiplegic migraine) between D19S215 and D19S216. These two diseases may be allelic disorders in which the same gene is involved [Tournier-Lasserre, E., et al., Nature
10 Genet., 3, 256-259 (1993); Joutel, A., et al., Nature Genet., 5, 40-45 (1993)].

[0105]

Although no evidence is available to support CKAP2 as a candidate gene for FHM or CADASIL, it is
15 conceivable that its mutation might lead to some or other neurological disease.

[0106]

By using the novel human CKAP2 gene of the present invention as obtained in this example, it is
20 possible to detect the expression of said gene in various tissues or produce the human CKAP2 gene in the manner of genetic engineering. Through these, it becomes possible to analyze the functions of the human CKAP2 system or human CKAP2, which is involved in diverse activities
25 essential to cells, as mentioned above, to diagnose

various neurological diseases in which said system or gene is involved, for example familial migraine, and to screen out and evaluate a therapeutic or prophylactic drug therefor.

5 [0107]

[Example 3] OTK27 gene

(1) OTK27 gene cloning and DNA sequencing

As a result of sequence analysis of cDNA clones arbitrarily selected from a human fetal brain cDNA
10 library in the same manner as in Example 1-(1) and database searching, a cDNA clone, GEN-025F07, coding for a protein highly homologous to NHP2, a yeast nucleoprotein [Saccharomyces cerevisiae; Kolodrubetz, D. and Burgum, A., YEAST, 7, 79-90 (1991)], was found and
15 named OTK27.

[0108]

Nucleoproteins are fundamental cellular constituents of chromosomes, ribosomes and so forth and are thought to play an essential role in cell
20 multiplication and viability. The yeast nucleoprotein NHP2, a high-mobility group (HMG)-like protein, like HMG, has reportedly a function essential for cell viability [Kolodrubetz, D. and Burgum, A., YEAST, 7, 79-90 (1991)].

[0109]

25 The novel human gene, OTK27 gene, of the

present invention, which is highly homologous to the above-mentioned yeast NHP2 gene, is supposed to be similar in function.

[0110]

5 The nucleotide sequence of said GEN-025F07 clone was found to comprise 1493 nucleotides, as shown under SEQ ID NO:9, and contain an open reading frame comprising 384 nucleotides, as shown under SEQ ID NO:8, coding for an amino acid sequence comprising 128 amino
10 acid residues, as shown under SEQ ID NO:7. The initiation codon was located at nucleotides 95-97 of the sequence shown under SEQ ID NO:9, and the termination codon at nucleotides 479-481.

[0111]

15 At the amino acid level, the OTK27 protein was highly homologous (38%) to NHP2. It was 83% identical with the protein deduced from the cDNA from Arabidopsis thaliana; Newman, T., unpublished; GENEMBL Accession No. T14197).

20 [0112]

(2) Northern blot analysis

For examining the expression of human OTK27 mRNA in normal human tissues, the insert in the OTK27 cDNA was amplified by PCR, the PCR product was purified
25 and labeled with [³²P]-dCTP (random-primed DNA labeling

kit, Boehringer Mannheim), and Northern blotting was performed using the labeled product as a probe in the same manner as in Example 1-(2).

[0113]

5 As a result of the Northern blot analysis, two bands corresponding to possible transcripts from this gene were detected at approximately 1.6 kb and 0.7 kb. Both sizes of transcript were expressed in all normal adult tissues examined. However, the expression of the
10 0.7 kb transcript was significantly reduced in brain and was of higher levels in heart, skeletal muscle and testicle than in other tissues examined.

[0114]

 For further examination of these two
15 transcripts, eleven cDNA clones were isolated from a testis cDNA library and their DNA sequences were determined in the same manner as in Example 1-(1).

[0115]

 As a result, in six clones, the sequences were
20 found to be in agreement with that of the 0.7 kb transcript, with a poly(A) sequence starting at around the 600th nucleotide, namely at the 598th nucleotide in two of the six clones, at the 606th nucleotide in three clones, and at the 613th nucleotide in one clone.

25 [0116]

In these six clones, the "TATAAA" sequence was recognized at nucleotides 583-588 as a probable poly(A) signal. The upstream poly(A) signal "TATAAA" of this gene was recognized as little influencing in brain and
5 more effective in the three tissues mentioned above than in other tissues. The possibility was considered that the stability of each transcript vary from tissue to tissue.

[0117]

10 Results of zoo blot analysis indicated that this gene is well conserved also in other vertebrates. Since this gene is expressed ubiquitously in normal adult tissues and conserved among a wide range of species, the gene product is likely to play an important physiological
15 role. The evidence that yeasts lacking in NHP2 are nonviable suggests that the human homolog may also be essential to cell viability.

[0118]

(3) Chromosomal localization of OTK27 by direct R-
20 banding FISH

One cosmid clone corresponding to the cDNA OTK27 was isolated from a total human genomic cosmid library (5-genome equivalent) using the OTK27 cDNA insert as a probe and subjected to FISH in the same manner as in
25 Example 1-(3) for chromosomal localization of OTK27.

[0119]

As a result, two distinct spots were observed on the chromosome band 12q24.3.

[0120]

5 The OTK27 gene of the present invention can be used in causing expression thereof and detecting the OTK27 protein, a human nucleoprotein, and thus can be utilized in the diagnosis and pathologic studies of various diseases in which said protein is involved and,
10 because of its involvement in cell proliferation and differentiation, in screening out and evaluating therapeutic and preventive drugs for cancer.

[0121]

[Example 4] OTK18 gene

15 (1) OTK18 gene cloning and DNA sequencing

 Zinc finger proteins are defined as constituting a large family of transcription-regulating proteins in eukaryotes and carry evolutionally conserved structural motifs [Kadonaga, J. T., et al., Cell, 51,
20 1079-1090 (1987); Klung, A. and Rhodes, D., Trends Biol. Sci., 12, 464-469 (1987); Evans, R. M. and Hollenberg, S. M., Cell, 52, 1-3 (1988)].

[0122]

 The zinc finger, a loop-like motif formed by
25 the interaction between the zinc ion and two residues,

cysteine and histidine residues, is involved in the sequence-specific binding of a protein to RNA or DNA. The zinc finger motif was first identified within the amino acid sequence of the Xenopus transcription factor IIIA [Miller, J., et al., EMBO J., 4, 1609-1614 (1986)].

5 [0123]

The C₂H₂ finger motif is in general tandemly repeated and contains an evolutionally conserved intervening sequence of 7 or 8 amino acids. This intervening stretch was first identified in the Kruppel segmentation gene of Drosophila [Rosenberg, U. B., et al., Nature, 319, 336-339 (1986)]. Since then, hundreds of C₂H₂ zinc finger protein-encoding genes have been found in vertebrate genomes.

10

15 [0124]

As a result of sequence analysis of cDNA clones arbitrarily selected from a human fetal brain cDNA library in the same manner as in Example 1-(1) and database searching, several zinc finger structure-containing clones were identified and, further, a clone

20

having a zinc finger structure of the Kruppel type was found.

[0125]

Since this clone lacked the 5' portion of the transcript, plaque hybridization was performed with a

25

fetal brain cDNA library using, as a probe, an approximately 1.8 kb insert in the cDNA clone, whereby three clones were isolated. The nucleotide sequences of these were determined in the same manner as in Example 1-(1).

5 [0126]

Among the three clones, the one having the largest insert spans 3,754 nucleotides including an open reading frame of 2,133 nucleotides coding for 711 amino acids. It was found that said clone contains a novel
10 human gene coding for a peptide highly homologous in the zinc finger domain to those encoded by human ZNF41 and the Drosophila Kruppel gene. This gene was named OTK18 gene (derived from the clone GEN-076C09).

[0127]

15 The nucleotide sequence of the cDNA clone of the OTK18 gene is shown under SEQ ID NO:12, the coding region-containing nucleotide sequence under SEQ ID NO:11, and the predicted amino acid sequence encoded by said OTK18 gene under SEQ ID NO:10.

20 [0128]

It was found that the amino acid sequence of OTK18 as deduced from SEQ ID NO:12 contains 13 finger motifs on its carboxy side.

[0129]

25 (2) Comparison with other zinc finger motif-containing

genes

Comparison among OTK18, human ZNF41 and the Drosophila Kruppel gene revealed that each finger motif is for the most part conserved in the consensus sequence

5 CXECGKAFFXQKSXLX₂HQRXH.

[0130]

Comparison of the consensus sequence of the zinc finger motifs of OTK18 with those of human ZNF41 and the Drosophila Kruppel gene revealed that the Kruppel
10 type motif is well conserved in the OTK18-encoded protein. However, the sequence similarities were limited to zinc finger domains and no significant homologies were found with regard to other regions.

[0131]

15 The zinc finger domain interacts specifically with the target DNA, recognizing an about 5 bp sequence to thereby bind to the DNA helix [Rhodes, D. and Klug, A., Cell, 46, 123-132 (1986)].

[0132]

20 Based on the idea that, in view of the above, the multiple module (tandem repetitions of zinc finger) can interact with long stretches of DNA, it is presumable that the target DNA of this gene product containing 13 repeated zinc finger units would be a DNA fragment with a
25 length of approximately 65 bp.

[0133]

(3) Northern blot analysis

Northern blot analysis was performed as described in Example 1-(2) for checking normal human tissues for expression of the human OTK18 mRNA therein by amplifying the insert of the OTK18 cDNA by PCR, purifying the PCR product, labeling the same with [32 P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim) and using an MTN blot with the labeled product as a probe.

[0134]

The results of Northern blot analysis revealed that the transcript of OTK18 is approximately 4.3 kb long and is expressed ubiquitously in various normal adult tissues. However, the expression level in the liver and in peripheral blood lymphocytes seemed to be lower than in other organs tested.

[0135]

(4) Cosmid cloning and chromosomal localization by direct R-banding FISH

Chromosomal localization of OTK18 was carried out as described in Example 1-(3).

[0136]

As a result, complete twin spots were identified with 8 samples while 23 samples showed an incomplete signal or twin spots on either or both

homologs. All signals appeared at the q13.4 band of chromosome 19. No twin spots were observed on any other chromosomes.

[0137]

5 The results of FISH thus revealed that this gene is localized on chromosomal band 19q13.4. This region is known to contain many DNA segments that hybridize with oligonucleotides corresponding to zinc finger domains [Hoovers, J. M. N., et al., Genomics, 12,
10 254-263 (1992)]. In addition, at least one other gene coding for a zinc finger domain has been identified in this region [Marine, J.-C., et al., Genomics, 21, 285-286 (1994)].

[0138]

15 Hence, the chromosome 19q13 is presumably a site of grouping of multiple genes coding for transcription-regulating proteins.

[0139]

20 When the novel human OTK18 gene provided by this example is used, it becomes possible to detect expression of said gene in various tissues and produce the human OTK18 protein in the manner of genetic engineering. Through these, it is possible to analyze the functions of the human transcription regulating
25 protein gene system or human transcription regulating

proteins, which are deeply involved in diverse activities
fundamental to cells, as mentioned above, to diagnose
various diseases with which said gene is associated, for
example malformation or cancer resulting from a
5 developmental or differentiation anomaly, and mental or
nervous disorder resulting from a developmental anomaly
in the nervous system, and further to screen out and
evaluate therapeutic or prophylactic drugs for these
diseases.

10 [0140]

[Example 5] Genes encoding human 26S proteasome
constituent P42 protein and P27 protein

(1) Cloning and DNA sequencing of genes respectively
encoding human 26S proteasome constituent P42
15 protein and P27 protein

Proteasome, which is a multifunctional
protease, is an enzyme occurring widely in eukaryotes
from yeasts to humans and decomposing ubiquitin-binding
proteins in cells in an energy-dependent manner.

20 Structurally, said proteasome is constituted of 20S
proteasome composed of various constituents with a
molecular weight of 21 to 31 kilodaltons and a group of
PA700 regulatory proteins composed of various
constituents with a molecular weight of 30 to 112
25 kilodaltons and showing a sedimentation coefficient of

22S and, as a whole, occurs as a macromolecule with a
molecular weight of about 2 million daltons and a
sedimentation coefficient of 26S [Rechsteiner, M., et
al., J. Biol. Chem., 268, 6065-6068 (1993); Yoshimura,
5 T., et al., J. Struct. Biol., 111, 200-211 (1993);
Tanaka, K., et al., New Biologist, 4, 173-187 (1992)].

[0141]

Despite structural and mechanical analyses
thereof, the whole picture of proteasome is not yet fully
10 clear. However, according to studies using yeasts and
mice in the main, it reportedly has the functions
mentioned below and its functions are becoming more and
more elucidated.

[0142]

15 The mechanism of energy-dependent proteolysis
in cells starts with selection of proteins by ubiquitin
binding. It is not 20S proteasome but 26S proteasome
that has ubiquitin-conjugated protein decomposing
activity which is ATP-dependent [Chu-Ping et al., J.
20 Biol. Chem., 269, 3539-3547 (1994)]. Hence, human 26S
proteasome is considered to be useful in elucidating the
mechanism of energy-dependent proteolysis.

[0143]

Factors involved in the cell cycle regulation
25 are generally short in half-life and in many cases they

are subject to strict quantitative control. In fact, it has been made clear that the oncogene products Mos, Myc, Fos and so forth can be decomposed by 26S proteasome in an energy- and ubiquitin-dependent manner [Ishida, N., et al., FEBS Lett., 324, 345-348 (1993); Hershko, A. and Ciechanover, A., Annu. Rev. Biochem., 61, 761-807 (1992)] and the importance of proteasome in cell cycle control is being recognized.

[0144]

10 Its importance in the immune system has also been pointed out. It is suggested that proteasome is positively involved in class I major histocompatible complex antigen presentation [Michalek, M. T., et al., Nature, 363, 552-554 (1993)] and it is further suggested
15 that proteasome may be involved in Alzheimer disease, since the phenomena of abnormal accumulation of ubiquitin-conjugated proteins in the brain of patients with Alzheimer disease [Kitaguchi, N., et al., Nature, 361, 530-532 (1988)]. Because of its diverse functions
20 such as those mentioned above, proteasome attracts attention from the viewpoint of its utility in the diagnosis and treatment of various diseases.

[0145]

25 A main function of 26S proteasome is ubiquitin-conjugated protein decomposing activity. In particular,

it is known that cell cycle-related gene products such as oncogene products and cyclins, typically c-Myc, are degraded via ubiquitin-dependent pathways. It has also been observed that the proteasome gene is expressed
5 abnormally in liver cancer cells, renal cancer cells, leukemia cells and the like as compared with normal cells [Kanayama, H., et al., Cancer Res., 51, 6677-6685 (1991)] and that proteasome is abnormally accumulated in tumor cell nuclei. Hence, constituents of proteasome are
10 expected to be useful in studying the mechanism of such canceration and in the diagnosis or treatment of cancer.

[0146]

Also, it is known that the expression of proteasome is induced by interferon γ and so on and is
15 deeply involved in antigen presentation in cells [Aki, M., et al., J. Biochem., 115, 257-269 (1994)]. Hence, constituents of human proteasome are expected to be useful in studying the mechanism of antigen presentation in the immune system and in developing immunoregulating
20 drugs.

[0147]

Furthermore, proteasome is considered to be deeply associated with ubiquitin abnormally accumulated in the brain of patients with Alzheimer disease. Hence,
25 it is suggested that constituents of human proteasome

should be useful in studying the cause of Alzheimer disease and in the treatment of said disease.

[0148]

In addition to the utilization of expectedly
5 multifunctional proteasome as such in the above manner,
it is probably possible to produce antibodies using
constituents of proteasome as antigens and use such
antibodies in diagnosing various diseases by immunoassay.
Its utility in this field of diagnosis is thus also a
10 focus of interest.

[0149]

Meanwhile, a protein having the characteristics
of human 26S proteasome is disclosed, for example in
Japanese Unexamined Patent Publication No. 292964/1993
15 and rat proteasome constituents are disclosed in Japanese
Unexamined Patent Publication Nos. 268957/1993 and
317059/1993. However, no human 26S proteasome
constituents are known. Therefore, the present inventors
made a further search for human 26S proteasome
20 constituents and successfully obtained two novel human
26S proteasome constituents, namely human 26S proteasome
constituent P42 protein and human S26 proteasome
constituent P27 protein, and performed cloning and DNA
sequencing of the corresponding genes in the following
25 manner.

[0150]

(1) Purification of human 26S proteasome constituents
P42 protein and P27 protein

Human proteasome was purified using about 100 g
5 of fresh human kidney and following the method of purifying human proteasome as described in Japanese Unexamined Patent Publication No. 292964/1993, namely by column chromatography using BioGel A-1.5 m (5 x 90 cm, Bio-Rad), hydroxyapatite (1.5 x 15 cm, Bio-Rad) and Q-Sepharose
10 (1.5 x 15 cm, Pharmacia) and glycerol density gradient centrifugation.

[0151]

The thus-obtained human proteasome was subjected to reversed phase high performance liquid
15 chromatography (HPLC) using a Hitachi model L6200 HPLC system. A Shodex RS Pak D4-613 (0.6 x 15 cm, Showa Denko) was used and gradient elution was performed with the following two solutions:

[0152]

20 First solution: 0.06% trifluoroacetic acid;
Second solution: 0.05% trifluoroacetic acid, 70% acetonitrile.

An aliquot of each eluate fraction was subjected to 8.5% SDS-polyacrylamide electrophoresis
25 under conditions of reduction with dithiothreitol. The

P42 protein and P27 protein thus detected were isolated and purified.

[0153]

The purified P42 and P27 proteins were respectively digested with 1 μ g of trypsin in 0.1 M Tris buffer (pH 7.8) containing 2 M urea at 37°C for 8 hours and the partial peptide fragments obtained were separated by reversed phase HPLC and their sequences were determined by Edman degradation. The results obtained are as shown below in Table 2.

[0154]

[Table 2]

Partial protein	Amino acid sequence
P42 (1)	VLNISLW
(2)	TLMELLNQMDGFDTLHR
(3)	AVSDFVSEYXMXA
(4)	EVDPLVYNX
(5)	HGEIDYEAIVK
(6)	LSXGFNGADLRNVXTEAGMFAIXAD
(7)	MIMATNRPDTLDPALLRPGXL
(8)	IHIDLPNEQARLDILK
(9)	ATNGPRYVVVG
(10)	EIDGRLK
(11)	ALQSVGQIVGEVLK
(12)	ILAGPITK
(13)	XXVIELPLTNPELFQG
(14)	VVSSSLVDK
(15)	ALQDYRK
(16)	EHREQLK
(17)	KLESKLDYKPVR
P27 (1)	LVPTR
(2)	AKEEEIEAQIK
(3)	ANYEVLESQK
(4)	VEDALHQLHAR
(5)	DVDLYQVR
(6)	QSQGLSPAQAFK
(7)	AGSQSGGSPEASGVTVSDVQE
(8)	GLLGXNI I PLQR

[0155]

(2) cDNA library screening, clone isolation and cDNA
nucleotide sequence determination

As mentioned in Example 1-(1), the present
5 inventors have a database comprising about 30,000 cDNA
data as constructed based on large-scale DNA sequencing
using human fetal brain, arterial blood vessel and
placenta cDNA libraries.

[0156]

10 Based on the amino acid sequences obtained as
mentioned above in (1), computer searching was performed
with the FASTA program (search for homology between said
amino acid sequences and the amino acid sequences
estimated from the database). As regards P42, a clone
15 (GEN-331G07) showing identity with regard to two amino
acid sequences [(2) and (7) shown in Table 2] was
screened out and, as regards P27, a clone (GEN-163D09)
showing identity with regard to two amino acid sequences
[(1) and (8) shown in Table 2] was found.

20 [0157]

For each of these clones, the 5' side sequence
was determined by 5' RACE and the whole sequence was
determined, in the same manner as in Example 2-(2).

[0158]

25 As a result, it was revealed that the above-

mentioned P42 clone GEN-331G07 comprises a 1,566-nucleotide sequence as shown under SEQ ID NO:15, inclusive of a 1,167-nucleotide open reading frame as shown under SEQ ID NO:14, and that the amino acid
5 sequence encoded thereby is the one shown under SEQ ID NO:13 and comprises 389 amino acid residues.

[0159]

The results of computer homology search revealed that the P42 protein is significantly homologous
10 to the AAA (ATPase associated with a variety of cellular activities) protein family (e.g. P45, TBP1, TBP7, S4, MSS1, etc.). It was thus suggested that it is a new member of the AAA protein family.

[0160]

As for the P27 clone GEN-163D09, it was revealed that it comprises a 1,128-nucleotide sequence as shown under SEQ ID NO:18, including a 669-nucleotide open reading frame as shown under SEQ ID NO:17 and that the amino acid sequence encoded thereby is the one shown
20 under SEQ ID NO:16 and comprises 223 amino acid residues.

[0161]

As regards the P27 protein, homology search using a computer failed to reveal any homologous gene among public databases. Thus, the gene in question is
25 presumably a novel gene having an unknown function.

[0162]

Originally, the above-mentioned P42 and P27 gene products were both purified as regulatory subunit components of proteasome complex. Therefore, these are
5 expected to play an important role in various biological functions through proteolysis, for example a role in energy supply through decomposition of ATP and, hence, they are presumably useful not only in studying the function of human 26S proteasome but also in the
10 diagnosis and treatment of various diseases caused by lowering of said biological functions, among others.

[0163]

[Example 6] BNAP gene

(1) BNAP gene cloning and DNA sequencing

15 The nucleosome composed of DNA and histone is a fundamental structure constituting chromosomes in eukaryotic cells and is well conserved over borders among species. This structure is closely associated with the processes of replication and transcription of DNA.
20 However, the nucleosome formation is not fully understood as yet. Only certain specific factors involved in nucleosome assembly (NAPs) have been identified. Thus, two acidic proteins, nucleoplasmin and N1, are already known to facilitate nucleosome construction
25 [Kleinschmidt, J. A., et al., J. Biol. Chem., 260, 1166-

1176 (1985); Dilworth, S. M., et al., Cell, 51, 1009-1018 (1987)].

[0164]

A yeast gene, NAP-I, was isolated using a monoclonal antibody and recombinant proteins derived therefrom were tested as to whether they have nucleosome assembling activity in vitro.

[0165]

More recently, a mouse NAP-I gene, which is a mammalian homolog of the yeast NAP-I gene was cloned (Okuda, A.; registered in database under the accession number D12618). Also cloned were a mouse gene, DN38 [Kato, K., Eur. J. Neurosci., 2, 704-711 (1990)] and a human nucleosome assembly protein (hNRP) [Simon, H. U., et al., Biochem. J., 297, 389-397 (1994)]. It was shown that the hNRP gene is expressed in many tissues and is associated with T lymphocyte proliferation.

[0166]

The present inventors performed sequence analysis of cDNA clones arbitrarily chosen from a human fetal brain cDNA library in the same manner as in Example 1-(1), followed by searches among databases and, as a result, made it clear that a 1,125-nucleotide cDNA clone (free of poly(A)), GEN-078D05, is significantly homologous to the mouse NAP-I gene, which is a gene for a

nucleosome assembly protein (NAP) involved in nucleosome construction, a mouse partial cDNA clone, DN38, and hNRP.

[0167]

Since said clone GEN-078D05 was lacking in the
5 5' region, 5' RACE was performed in the same manner as in
Example 2-(2) to obtain the whole coding region. For
this 5' RACE, primers P1 and P2 respectively having the
nucleotide sequences shown below in Table 3.

[0168]

10 [Table 3]

Primer		Nucleotide sequence
Primer P1	5'	-TTGAAGAATGATGCATTAGGAACCAC-3'
Primer P2	5'	-CACTCGAGTGGCTGGATTTCAATTTCTCCAGTAG-3'

[0169]

After the first 5' RACE, a single band
20 corresponding to a sequence length of 1,300 nucleotides
was obtained. This product was inserted into pT7Blue(R)
T-Vector and several clones appropriate in insert size
were selected.

[0170]

25 Ten 5' RACE clones obtained from two
independent PCR reactions were sequenced and the longest
clone GEN-078D05TA13 (about 1,300 nucleotides long) was
further analyzed.

[0171]

Both strands of the two overlapping cDNA clones GEN-078D05 and GEN-078D05TA13 were sequenced, whereby it was confirmed that the two clones did not yet cover the whole coding region. Therefore, a further second 5' RACE was carried out. For the second 5' RACE, two primers, P3 and P4, respectively having the sequences shown below in Table 4 were used.

[0172]

10 [Table 4]

Primer	Nucleotide sequence
Primer P3	5'-GTCGAGCTAGCCATCTCCTCTTCG-3'
Primer P4	5'-CATGGGCGACAGGTTCCGAGACC-3'

[0173]

A clone, GEN-078D0508, obtained by the second 5' RACE was 300 nucleotides long. This clone contained an estimable initiation codon and three preceding in-frame termination codons. From these three overlapping clones, it became clear that the whole coding region comprises 2,636 nucleotides. This gene was named brain-specific nucleosome assembly protein (BNAP) gene.

[0174]

The BNAP gene contains a 1,518-nucleotide open reading frame shown under SEQ ID NO:20. The amino acid

encoded thereby comprises 506 amino acid residues, as shown under SEQ ID NO:19, and the nucleotide sequence of the whole cDNA clone of BNAP is as shown under SEQ ID NO:21.

5 [0175]

As shown under SEQ ID NO:21, the 5' noncoding region of said gene was found to be generally rich in GC. Candidate initiation codon sequences were found at nucleotides Nos. 266-268, 287-289 and 329-331. These
10 three sequences all had well conserved sequences in the vicinity of the initiation codons [Kozak, M., J. Biol. Chem., 266, 19867-19870 (1991)].

[0176]

According to the scanning model, the first ATG
15 (nucleotides Nos. 266-268) of the cDNA clone may be the initiation codon. The termination codon was located at nucleotides Nos. 1784-1786.

[0177]

The 3' noncoding region was generally rich in
20 AT and two polyadenylation signals (AATAAA) were located at nucleotides Nos. 2606-2611 and 2610-2615, respectively.

[0178]

The longest open reading frame comprised 1,518
25 nucleotides coding for 506 amino acid residues and the

calculated molecular weight of the BNAP gene product was 57,600 daltons.

[0179]

Hydrophilic plots indicated that BNAP is very
5 hydrophilic, like other NAPs.

[0180]

For recombinant BNAP expression and purification and for eliminating the possibility that the BNAP gene sequence might give three chimera clones in the
10 step of 5' RACE, RT-PCR was performed using a sequence comprising nucleotides Nos. 326-356 as a sense primer and a sequence comprising nucleotides Nos. 1758-1786 as an antisenses primer.

[0181]

15 As a result, a single product of about 1,500 bp was obtained and it was thus confirmed that said sequence is not a chimera but a single transcript.

[0182]

(2) Comparison between BNAP and NAPs

20 The amino acid sequence deduced from BNAP showed 46% identity and 65% similarity to hNRP.

[0183]

The deduced BNAP gene product had motifs characteristic of the NAPs already reported and of BNAP.
25 In general, half of the C terminus was well conserved in

humans and yeasts.

[0184]

The first motif (domain I) is KGIPDYWLI
(corresponding to amino acid residues Nos. 309-317).

5 This was observed also in hNRP (KGIPSFWLT) and in yeast
NAP-I (KGIPEFWLT).

[0185]

The second motif (domain II) is ASFFNFFSPP
(corresponding to amino acid residues Nos. 437-446) and
10 this was expressed as DSFFNFFAPP in hNRP and as ESFFNFFSP
in yeast NAP-I.

[0186]

These two motifs were also conserved in the
deduced mouse NAP-I and DN38 peptides. Both conserved
15 motifs were each a hydrophilic cluster, and the Cys in
position 402 was also found conserved.

[0187]

Half of the N terminus had no motifs strictly
conserved from yeasts to mammalian species, while motifs
20 conserved among mammalian species were found.

[0188]

For instance, HDLERKYA (corresponding to amino
acid residues Nos. 130 to 137) and IINAEYEPTEEECEW
(corresponding to amino acid residues Nos. 150-164),
25 which may be associated with mammal-specific functions,

were found strictly conserved.

[0189]

NAPs had acidic stretches, which are believed to be readily capable of binding to histone or other
5 basic proteins. All NAPs had three acidic stretches but the locations thereof were not conserved.

[0190]

BNAP has no such three acidic stretches but, instead, three repeated sequences (corresponding to amino
10 acid residues Nos. 194-207, 208-221 and 222-235) with a long acidic cluster, inclusive of 41 amino acid residues out of 98 amino acid residues, the consensus sequence being ExxKExPEVKxEEK (each x being a nonconserved, mostly hydrophobic, residue).

15 [0191]

Furthermore, it was revealed that the BNAP sequence had several BNAP-specific motifs. Thus, an extremely serine-rich domain (corresponding to amino acid residues Nos. 24-72) with 33 (67%) of 49 amino acid
20 residues being serine residues was found in the N-terminus portion. On the nucleic acid level, they were reflected as incomplete repetitions of AGC.

[0192]

Following this serine-rich region, there
25 appeared a basic domain (corresponding to amino acid

residues Nos. 71-89) comprising 10 basic amino acid residues among 19 residues.

[0193]

BNAP is supposed to be localized in the nucleus. Two possible signals localized in the nucleus were observed (NLSs). The first signal was found in the basic domain of BNAP and its sequence YRKRR (corresponding to amino acid residues Nos. 75-79) was similar to NLS (GRKKR) of Tat of HIV-1. The second signal was located in the C terminus and its sequence KKYRK (corresponding to amino acid residues Nos. 502-506) was similar to NLS (KKKRR) of the large T antigen of SV40. The presence of these two presumable NLSs suggested the localization of BNAP in the nucleus. However the possibility that other basic clusters might act as NLSs could not be excluded.

[0194]

BNAP has several phosphorylation sites and the activity of BNAP may be controlled through phosphorylation thereof.

[0195]

(3) Northern blot analysis

Northern blot analysis was performed as described in Example 1-(2). Thus, the clone GEN-078D05TA13 (corresponding to nucleotides Nos. 323 to 1558

in the BNAP gene sequence) was amplified by PCR, the PCR product was purified and labeled with [^{32}P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim), and the expression of BNAP mRNA in normal human tissues was
5 examined using an MTN blot with the labeled product as a probe.

[0196]

As a result of Northern blot analysis, a 3.0 kb transcript of BNAP was detected (8-hour exposure) in the
10 brain among eight human adult tissues tested, namely heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas and, after longer exposure (24 hours), a dim band of the same size was detected in the heart.

15 [0197]

BNAP was found equally expressed in several sites of brain tested whereas, in other tissues, no signal was detected at all even after 72 hours of exposure. hNRP mRNA was found expressed everywhere in
20 the human tissues tested whereas the expression of BNAP mRNA was tissue-specific.

[0198]

(4) Radiation hybrid mapping

Chromosomal mapping of the BNAP clone was
25 performed by means of radiation hybrid mapping [Cox, D.

R., et al., Science, 250, 245-250 (1990)].

[0199]

Thus, a total human genome radiation hybrid
clone (G3RH) panel was purchased from Research Genetics,
5 Inc., AL, USA and PCR was carried out for chromosomal
mapping analysis according to the product manual using
two primers, A1 and A2, respectively having the
nucleotide sequences shown in Table 5.

[0200]

10 [Table 5]

Primer		Nucleotide sequence
A1 primer		5'-CCTAAAAAGTGTCTAAGTGCCAGTT-3'
A2 primer		5'-TCAGTGAAAGGGAAGGTAGAACAC-3'

15

[0201]

The results obtained were analyzed utilizing
20 softwares usable on the Internet [Boehnke, M., et al.,
Am. J. Hum. Genet., 46, 581-586 (1991)].

[0202]

As a result, the BNAP gene was found strongly
linked to the marker DXS990 (LOD = 1000, cR8000 = -0.00).
25 Since DXS990 is a marker localized on the chromosome
Xq21.3-q22, it was established that BNAP is localized to
the chromosomal locus Xq21.3-q22 where genes involved in
several signs or symptoms of X-chromosome-associated

mental retardation are localized.

[0203]

The nucleosome is not only a fundamental
chromosomal structural unit characteristic of eukaryotes
5 but also a gene expression regulating unit. Several
results indicate that genes with high transcription
activity are sensitive to nuclease treatment, suggesting
that the chromosome structure changes with the
transcription activity [Elgin, S. C. R., J. Biol. Chem.,
10 263, 19259-19262 (1988)].

[0204]

NAP-I has been cloned in yeast, mouse and human
and is one of the factors capable of promoting nucleosome
construction in vitro. In a study performed on their
15 sequences, NAPs containing the epitope of the specific
antibody 4A8 were detected in human, mouse, frog,
Drosophila and yeast (Saccharomyces cerevisiae) [Ishimi,
Y., et al., Eur. J. Biochem., 162, 19-24 (1987)].

[0205]

20 In these experiments, NAPs, upon SDS-PAGE
analysis, electrophoretically migrated to positions
corresponding to a molecular weight between 50 and 60
kDa, whereas the recombinant BNAP slowly migrated to a
position of about 80 kDa. The epitope of 4A8 was shown
25 to be localized in the second, well-conserved,

hydrophobic motif. And, it was simultaneously shown that the triplet FNF is important as a part of the epitope [Fujii-Nakata, T., et al., J. Biol. Chem., 267, 20980-20986 (1992)].

5 [0206]

BNAP also contained this consensus motif in domain II. The fact that domain II is markedly hydrophobic and the fact that domain II can be recognized by the immune system suggest that it is probably presented on the BNAP surface and is possibly involved in protein-protein interactions.

[0207]

Domain I, too, may be involved in protein-protein interactions. Considering that these are conserved generally among NAPs, though to a relatively low extent, it is conceivable that they must be essential for nucleosome construction, although the functional meaning of the conserved domains is still unknown.

[0208]

20 The hNRP gene is expressed in thyroid gland, stomach, kidney, intestine, leukemia, lung cancer, mammary cancer and so on [Simon, H. U., et al., Biochem. J., 297, 389-397 (1994)]. Like that, NAPs are expressed everywhere and are thought to be playing an important role in fundamental nucleosome formation.

25

[0209]

BNAP may be involved in brain-specific
nucleosome formation and an insufficiency thereof may
cause neurological diseases or mental retardation as a
5 result of deviated functions of neurons.

[0210]

BNAP was found strongly linked to a marker on
the X-chromosome q21.3-q22 where sequences involved in
several symptoms of X-chromosome-associated mental
10 retardation are localized. This center-surrounding
region of X-chromosome was rich in genes responsible for
 α -thalassemia, mental retardation (ATR-X) or some other
forms of mental retardation [Gibbons, R. J., et al.,
Cell, 80, 837-845 (1995)]. Like the analysis of the ATR-
15 X gene which seems to regulate the nucleosome structure,
the present inventors suppose that BNAP may be involved
in a certain type of X-chromosome-linked mental
retardation.

[0211]

20 According to this example, the novel BNAP gene
is provided and, when said gene is used, it is possible
to detect the expression of said gene in various tissues
and to produce the BNAP protein by the technology of
genetic engineering. Through these, it is possible to
25 study the brain nucleosome formation deeply involved, as

mentioned above, in variegated activities essential to cells as well as the functions of cranial nerve cells and to diagnose various neurological diseases or mental retardation in which these are involved and screen out
5 and evaluate drugs for the treatment or prevention of such diseases.

[0212]

[Example 7] Human skeletal muscle-specific ubiquitin-conjugating enzyme gene (UBE2G gene)

10 The ubiquitin system is a group of enzymes essential for cellular processes and is conserved from yeast to human. Said system is composed of ubiquitin-activating enzymes (UBAs), ubiquitin-conjugating enzymes (UBCs), ubiquitin protein ligases (UBRs) and 26S
15 proteasome particles.

[0213]

Ubiquitin is transferred from the above-mentioned UBAs to several UBCs, whereby it is activated. UBCs transfer ubiquitins to target proteins with or
20 without the participation of UBRs. These ubiquitin-conjugated target proteins are said to induce a number of cellular responses, such as protein degradation, protein modification, protein translocation, DNA repair, cell cycle control, transcription control, stress responses,
25 etc. and immunological responses [Jentsch, S., et al.,

Biochim. Biophys. Acta, 1089, 127-139 (1991); Hershko, A. and Ciechanover, A., Annu. Rev. Biochem., 61, 761-807 (1992); Jentsch, S., Annu. Rev. Genet., 26, 179-207 (1992); Ciechanover, A., Cell, 72, 13-21 (1994)].

5 [0214]

UBCs are key components of this system and seem to have distinct substrate specificities and modulate different functions. For example, Saccharomyces cerevisiae UBC7 is induced by cadmium and involved in resistance to cadmium poisoning [Jungmann, J., et al., Nature, 361, 369-371 (1993)]. Degradation of MAT- α 2 is also executed by UBC7 and UBC6 [Chen, P., et al., Cell, 74, 357-369 (1993)].

[0215]

15 The novel gene obtained in this example is UBC7-like gene strongly expressed in human skeletal muscle. In the following, cloning and DNA sequencing thereof are described.

[0216]

20 (1) Cloning and DNA sequencing of human skeletal muscle-specific ubiquitin-conjugating enzyme gene (UBE2G gene)

Following the same procedure as in Example 1-(1), cDNA clones were arbitrarily selected from a human fetal brain cDNA library and subjected to sequence

25

analysis, and database searches were performed. As a
result, a cDNA clone, GEN-423A12, was found to have a
significantly high level of homology to the genes coding
for ubiquitin-conjugating enzymes (UBCs) in various
5 species.

[0217]

Since said GEN-423A12 clone was lacking in the
5' side, 5' RACE was performed in the same manner as in
Example 2-(2) to obtain an entire coding region.

10 [0218]

For said 5' RACE, two primers, P1 and P2,
respectively having the nucleotide sequences shown in
Table 6 were used.

[0219]

15 [Table 6]

Primer	Nucleotide sequence
P1 primer	5'-TAATGAATTTCATTTTAGGAGGTCGG-3'
P2 primer	5'-ATCTTTTGGGAAAGTAAGATGAGCC-3'

20

[0220]

The 5' RACE product was inserted into
25 pT7Blue(R) T-Vector and clones with an insert proper in
size were selected.

[0221]

Four of the 5' RACE clones obtained from two

independent PCR reactions contained the same sequence but were different in length.

[0222]

By sequencing the above clones, the coding
5 sequence and adjacent 5'- and 3'-flanking sequences of the novel gene were determined.

[0223]

As a result, it was revealed that the novel gene has a total length of 617 nucleotides. This gene
10 was named human skeletal muscle-specific ubiquitin-conjugating enzyme gene (UBE2G gene).

[0224]

To exclude the conceivable possibility that this sequence was a chimera clone, RT-PCR was performed
15 in the same manner as in Example 6 (1) using the sense primer to amplify said sequence from the human fetal brain cDNA library. As a result, a single PCR product was obtained, whereby it was confirmed that said sequence is not a chimera one.

20 [0225]

The UBE2G gene contains an open reading frame of 510 nucleotides, which is shown under SEQ ID NO:23, the amino acid sequence encoded thereby comprises 170 amino acid residues, as shown under SEQ ID NO:22, and the
25 nucleotide sequence of the entire UBE2G cDNA is as shown

under SEQ ID NO:24.

[0226]

As shown under SEQ ID NO:24, the estimable initiation codon was located at nucleotides Nos. 19-21, corresponding to the first ATG triplet of the cDNA clone. Since no preceding in-frame termination codon was found, it was deduced that this clone contains the entire open reading frame on the following grounds.

[0227]

Thus, (a) the amino acid sequence is highly homologous to S. cerevisiae UBC7 and said initiation codon agrees with that of yeast UBC7, supporting said ATG as such.

[0228]

(b) The sequence AGGATGA is similar to the consensus sequence (A/G)CCATGG around the initiation codon [Kozak, M., J. Biol. Chem., 266, 19867-19870 (1991)].

[0229]

(2) Comparison in amino acid sequence between UBE2G and UBCs

Comparison in amino acid sequence between UBE2G and UBCs suggested that the active site cysteine capable of binding to ubiquitin should be the 90th residue cysteine. The peptides encoded by these genes seem to belong to the same family.

[0230]

(3) Northern blot analysis

Northern blot analysis was carried out as described in Example 1-(2). Thus, the entire sequence of UBE2G was amplified by PCR, the PCR product was purified and labeled with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim) and the expression of UBE2G mRNA in normal human tissues using the labeled product as a probe. The membrane used was an MTN blot.

10 [0231]

As a result of the Northern blot analysis, 4.4 kb, 2.4 kb and 1.6 kb transcripts could be detected in all 16 human adult tissues, namely heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thyroid gland, urinary bladder, testis, ovary, small intestine, large intestine and peripheral blood leukocyte, after 18 hours of exposure. Strong expression of these transcripts was observed in skeletal muscle.

[0232]

20 (4) Radiation hybrid mapping

Chromosomal mapping of the UBE2G clone was performed by radiation hybrid mapping in the same manner as in Example 6-(4).

[0233]

25 The primers C1 and C4 used in PCR for

chromosomal mapping analysis respectively correspond to nucleotides Nos. 415-435 and nucleotides Nos. 509-528 in the sequence shown under SEQ ID NO:24 and their nucleotide sequences are as shown below in Table 7.

5 [0234]

[Table 7]

Primer	Nucleotide sequence
10 C1 primer	5'-GGAGACTCACCTGCTAATGTT-3'
C4 primer	5'-CTCAAAAGCAGTCTCTTGGC-3'

[0235]

15 As a result, the UBE2G gene was found linked to the markers D1S446 (LOD = 12.52, cR8000 = 8.60) and D1S235 (LOD = 9.14, cR8000 = 22.46). These markers are localized to the chromosome bands 1q42.13-q42.3.

[0236]

20 UBE2G was expressed strongly in skeletal muscle and very weakly in all other tissues examined. All other UBCs are involved in essential cellular functions, such as cell cycle control, and those UBCs are expressed ubiquitously. However, the expression pattern of UBE2G
25 might suggest a muscle-specific role thereof.

[0237]

While the three transcripts differing in size were detected, attempts failed to identify which

corresponds to the cDNA clone. The primary structure of the UBE2G product showed an extreme homology to yeast UBC7. On the other hand, nematode UBC7 showed strong homology to yeast UBC7. It is involved in degradation of the repressor and further confers resistance to cadmium in yeasts. The similarities among these proteins suggest that they belong to the same family.

[0238]

It is speculated that UBE2G is involved in degradation of muscle-specific proteins and that a defect in said gene could lead to such diseases as muscular dystrophy. Recently, another proteolytic enzyme, calpain 3, was found to be responsible for limb-girdle muscular dystrophy type 2A [Richard, I., et al., Cell, 81, 27-40 (1995)]. At the present, the chromosomal location of UBE2G suggests no significant relationship with any hereditary muscular disease but it is likely that a relation to the gene will be unearthed by linkage analysis in future.

[0239]

In accordance with this example, the novel UBE2G gene is provided and the use of said gene enables detection of its expression in various tissues and production of the UBE2G protein by the technology of genetic engineering. Through these, it becomes possible

to study the degradation of muscle-specific proteins
deeply involved in basic activities variegated and
essential to cells, as mentioned above, and the functions
of skeletal muscle, to diagnose various muscular diseases
5 in which these are involved and further to screen out and
evaluate drugs for the treatment and prevention of such
diseases.

[0240]

[Example 8] TMP-2 gene

10 (1) TMP-2 gene cloning and DNA sequencing

Following the procedure of Example 1-(1), cDNA
clones were arbitrarily selected from a human fetal brain
cDNA library and subjected to sequence analysis, and
database searches were performed. As a result, a clone
15 (GEN-092E10) having a cDNA sequence highly homologous to
a transmembrane protein gene (accession No.: U19878) was
found out.

[0241]

Membrane protein genes have so far been cloned
20 in frog (Xenopus laevis) and human. These are considered
to be a gene for a transmembrane type protein having a
follistatin module and an epidermal growth factor (EGF)
domain (accession No.: U19878).

[0242]

25 The sequence information of the above protein

gene indicated that the GEN-092E10 clone was lacking in the 5' region, so that the λ gt10 cDNA library (human fetal brain 5'-STRETCH PLUS cDNA; Clontech) was screened using the GEN-092E10 clone as a probe, whereby a cDNA
5 clone containing a further 5' upstream region was isolated.

[0243]

Both strands of this cDNA clone were sequenced, whereby the sequence covering the entire coding region
10 became clear. This gene was named TMP-2 gene.

[0244]

The TMP-2 gene was found to contain an open reading frame of 1,122 nucleotides, as shown under SEQ ID NO:26, encoding an amino acid sequence of 374 residues,
15 as shown under SEQ ID NO:25. The nucleotide sequence of the entire TMP-2 cDNA clone comprises 1,721 nucleotides, as shown under SEQ ID NO:27.

[0245]

As shown under SEQ ID NO:27, the 5' noncoding
20 region was generally rich in GC. Several candidates for the initiation codon were found but, according to the scanning model, the 5th ATG of the cDNA clone (bases Nos. 368-370) was estimated as the initiation codon. The termination codon was located at nucleotides Nos. 1490-
25 1492. The polyadenylation signal (AATAAA) was located at

nucleotides Nos. 1703-1708. The calculated molecular weight of the TMP-2 gene product was 41,400 daltons.

[0246]

As mentioned above, the transmembrane genes
5 have a follistatin module and an EGF domain. These motifs were also found conserved in the novel human gene of the present invention.

[0247]

The TMP-2 gene of the present invention
10 presumably plays an important role in cell proliferation or intercellular communication, since, on the amino acid level, said gene shows homology, across the EGF domain, to TGF- α (transforming growth factor- α ; Derynck, R., et al., Cell, 38, 287-297 (1984)), beta-cellulin [Igarashi,
15 K. and Folkman, J., Science, 259, 1604-1607 (1993)], heparin-binding EGF-like growth factor [Higashiyama, S., et al., Science, 251, 936-939 (1991)] and schwannoma-derived growth factor [Kimura, H., et al., Nature, 348, 257-260 (1990)].

20 [0248]

(2) Northern blot analysis

Northern blot analysis was carried out as described in Example 1-(2). Thus, the clone GEN-092E10 was amplified by PCR, the PCR product was purified and
25 labeled with [32 P]-dCTP (random-primed DNA labeling kit,

Boehringer Mannheim), and the expression of TMP-2 mRNA in normal human tissues was examined using an MTN blot with the labeled product as a probe.

[0249]

5 As a result, high levels of expression were detected in brain and prostate gland. Said TMP-2 gene mRNA was about 2 kb in size.

[0250]

 According to the present invention, the novel
10 human TMP-2 gene is provided and the use of said gene makes it possible to detect the expression of said gene in various tissues or produce the human TMP-2 protein by the technology of genetic engineering and, through these, it becomes possible to study brain tumor and prostatic
15 cancer, which are closely associated with cell proliferation or intercellular communication, as mentioned above, to diagnose these diseases and to screen out and evaluate drugs for the treatment and prevention of such diseases.

20 [0251]

[Example 9] Human NPIK gene

(1) Human NPIK gene cloning and DNA sequencing

 Following the procedures of Example 1 and Example 2, cDNA clones were arbitrarily selected from a
25 human fetal brain cDNA library and subjected to sequence

analysis, and database searches were performed. As a result, two cDNA clones highly homologous to the gene coding for an amino acid sequence conserved in phosphatidylinositol 3 and 4 kinases [Kunz, J., et al.,
5 Cell, 73, 585-596 (1993)] were obtained. These were named GEN-428B12c1 and GEN-428B12c2 and the entire sequences of these were determined as in the foregoing examples.

[0252]

10 As a result, the GEN-428B12c1 cDNA clone and the GEN-428B12c2 clone were found to have coding sequences differing by 12 amino acid residues at the 5' terminus, the GEN-428B12c1 cDNA clone being longer by 12 amino acid residues.

15 [0253]

The GEN-428B12c1 cDNA sequence of the human NPIK gene contained an open reading frame of 2,487 nucleotides, as shown under SEQ ID NO:32, encoding an amino acid sequence comprising 829 amino acid residues,
20 as shown under SEQ ID NO:31. The nucleotide sequence of the full-length cDNA clone comprised 3,324 nucleotides as shown under SEQ ID NO:33.

[0254]

The estimated initiation codon was located, as
25 shown under SEQ ID NO:33, at nucleotides Nos. 115-117

corresponding to the second ATG triplet of the cDNA clone. The termination codon was located at nucleotides Nos. 2602-2604 and the polyadenylation signal (AATAAA) at Nos. 3305-3310.

5 [0255]

On the other hand, the GEN-428B12c2 cDNA sequence of the human NPIK gene contained an open reading frame of 2,451 nucleotides, as shown under SEQ ID NO:29. The amino acid sequence encoded thereby comprised 817 amino acid residues, as shown under SEQ ID NO:28. The nucleotide sequence of the full-length cDNA clone comprised 3,602 nucleotides, as shown under SEQ ID NO:30.

[0256]

The estimated initiation codon was located, as shown under SEQ ID NO:30, at nucleotides Nos. 429-431 corresponding to the 7th ATG triplet of the cDNA clone. The termination codon was located at nucleotides Nos. 2880-2882 and the polyadenylation signal (AATAAA) at Nos. 3583-3588.

20 [0257]

(2) Northern blot analysis

Northern blot analysis was carried out as described in Example 1-(2). Thus, the entire sequence of human NPIK was amplified by PCR, the PCR product was purified and labeled with [³²P]-dCTP (random-primed DNA

labeling kit, Boehringer Mannheim), and normal human tissues were examined for expression of the human NPIK mRNA using the MTN blot membrane with the labeled product as a probe.

5 [0258]

As a result, the expression of the human NPIK gene was observed in 16 various human adult tissues examined and an about 3.8 kb transcript and an about 5 kb one could be detected.

10 [0259]

Using primer A having the nucleotide sequence shown below in Table 8 and containing the initiation codon of the GEN-428B12c2 cDNA and primer B shown in Table 8 and containing the termination codon, PCR was performed with Human Fetal Brain Marathon-Ready cDNA (Clontech) as a template, and the nucleotide sequence of the PCR product was determined.

[0260]

20 [Table 8]

Primer	Nucleotide sequence
Primer A	5'-ATGGGAGATACAGTAGTGGAGC-3'
Primer B	5'-TCACATGATGCCGTTGGTGAG-3'

[0261]

As a result, it was found that the human NPIK

mRNA expressed included one lacking in nucleotides Nos.
1060-1104 of the GEN-428B12c1 cDNA sequence (SEQ ID
NO:33) (amino acids Nos. 316-330 of the amino acid
sequence under SEQ ID NO:31) and one lacking in
5 nucleotides Nos. 1897-1911 of the GEN-428B12c1 cDNA
sequence (SEQ ID NO:33) (amino acids Nos. 595-599 of the
amino acid sequence under SEQ ID NO:31).

[0262]

It was further revealed that polymorphism
10 existed in this gene (428B12c1.fasta), as shown below in
Table 9, in the region of bases Nos. 1941-1966 of the
GEN-428B12c1 cDNA sequence shown under SEQ ID NO:33,
whereby a mutant protein was encoded which resulted from
the mutation of IQDSCEITT (amino acid residues Nos. 610-
15 618 in the amino acid sequence (SEQ ID NO:31) encoded by
GEN-428B12c1) into YKILVISA.

[0263]

[Table 9]

			1930	1940	1950	1959
			TGGATCAAGCCAATACAAGATTCTTGTA			
			TCCATTTGGGAACAGGAGCGAGTGCCCTTTGGATCAAGCC-ATACAAGATTCTTGTA--			
1900	1910	1920	1930	1940	1950	
1960	1970	1980				
ATTACGACTGATAGTGGCATG						
ATTCGGCTGATAGTGGCATGATTGAACCAAGTGGTCAATGCTGTGTCCATCCATCAGGTG						
1960	1970	1980	1990	2000	2010	

[0264]

(3) Chromosomal mapping of human NPIK gene by FISH

Chromosomal mapping of the human NPIK gene was carried out by FISH as described in Example 1-(3).

5 [0265]

As a result, it was found that the locus of the human NPIK gene is in the chromosomal position 1q21.1-q21.3.

[0266]

10 The human NPIK gene, a novel human gene, of the present invention included two cDNAs differing in the 5' region and capable of encoding 829 and 817 amino acid residues, as mentioned above. In view of this and further in view of the findings that the mRNA
15 corresponding to this gene includes two deletable sites and there occurs polymorphism in a specific region corresponding to amino acid residues Nos. 610-618 of the GEN-428B12c1 amino acid sequence (SEQ ID NO:31), whereby a mutant protein is encoded, it is conceivable that human
20 NPIK includes species resulting from a certain number of combinations, namely human NPIK, deletion-containing human NPIK, human NPIK mutant and/or deletion-containing human NPIK mutant.

25 Recently, several proteins belonging to the family including the above-mentioned PI3 and 4 kinases

have protein kinase activity [Dhand, R., et al., EMBO J.,
13, 522-533 (1994); Stack, J. H. and Emr, S. D., J. Biol.
Chem., 269, 31552-31562 (1994); Hartley, K. O., et al.,
Cell, 82, 848-856 (1995)].

5 [0267]

It was also revealed that a protein belonging
to this family is involved in DNA repair [Hartley, K. O.,
et al., Cell, 82, 849-856 (1995)] and is a causative gene
of ataxia [Savitsky, K., et al., Science, 268, 1749-1753
10 (1995)].

[0268]

It can be anticipated that the human NPIK gene-
encoded protein highly homologous to the family of these
PI kinases is a novel enzyme phosphorylating lipids or
15 proteins.

[0269]

According to this example, the novel human NPIK
gene is provided. The use of said gene makes it possible
to detect the expression of said gene in various tissues
20 and manufacture the human NPIK protein by the technology
of genetic engineering and, through these, it becomes
possible to study lipid- or protein-phosphorylating
enzymes such as mentioned above, study DNA repairing,
study or diagnose diseases in which these are involved,
25 for example cancer, and screen out and evaluate drugs for

the treatment or prevention thereof.

[0270]

[Example 10] nel-related protein type 1 (NRP1) gene and
nel-related protein type 2 (NRP2) gene

- 5 (1) Cloning and DNA sequencing of NRP1 gene and NRP2
gene

EGF-like repeats have been found in many
membrane proteins and in proteins related to growth
regulation and differentiation. This motif seems to be
10 involved in protein-protein interactions.

[0271]

Recently, a gene encoding nel, a novel peptide
containing five EGF-like repeats, was cloned from a chick
embryonic cDNA library [Matsushashi, S., et al., Dev.
15 Dynamics, 203, 212-222 (1995)]. This product is
considered to be a transmembrane molecule with its EGF-
like repeats in the extracellular domain. A 4.5 kb
transcript (nel mRNA) is expressed in various tissues at
the embryonic stage and exclusively in brain and retina
20 after hatching.

[0272]

Following the procedure of Example 1-(1), cDNA
clones were randomly selected from a human fetal brain
cDNA library and subjected to sequence analysis, followed
25 by database searching. As a result, two cDNA clones with

significantly high homology to the above-mentioned nel
were found and named GEN-073E07 and GEN-093E05,
respectively.

[0273]

5 Since both clones were lacking in the 5'
portion, 5' RACE was performed in the same manner as in
Example 2-(2) to obtain the entire coding regions.

[0274]

10 As for the primers for 5' RACE, primers having
an arbitrary sequence obtained from the cDNA sequences of
the above clones were synthesized while the anchor primer
attached to a commercial kit was used as such.

[0275]

15 5' RACE clones obtained from the PCR were
sequenced and the sequences seemingly covering the entire
coding regions of both genes were obtained. These genes
were respectively named nel-related protein type 1 (NRP1)
gene and nel-related protein type 2 (NRP2) gene.

[0276]

20 The NRP1 gene contains an open reading frame of
2,430 nucleotides, as shown under SEQ ID NO:35, the amino
acid sequence deduced therefrom comprises 810 amino acid
residues, as shown under SEQ ID NO:34, and the nucleotide
sequence of the entire cDNA clone of said NRP1 gene
25 comprises 2,977 nucleotides, as shown under SEQ ID NO:36.

[0277]

On the other hand, the NRP2 gene contains an open reading frame of 2,448 nucleotides, as shown under SEQ ID NO:38, the amino acid sequence deduced therefrom
5 comprises 816 amino acid residues, as shown under SEQ ID NO:37, and the nucleotide sequence of the entire cDNA clone of said NRP2 gene comprises 3,198 nucleotides, as shown under SEQ ID NO:39.

[0278]

10 Furthermore, the coding regions were amplified by RT-PCR to exclude the possibility that either of the sequences obtained was a chimeric cDNA.

[0279]

The deduced NRP1 and NRP2 gene products both
15 showed highly hydrophobic N termini capable of functioning as signal peptides for membrane insertion. As compared with chick embryonic nel, they both appeared to have no hydrophobic transmembrane domain. Comparison among NRP1, NRP2 and nel with respect to the deduced
20 peptide sequences revealed that NRP2 has 80% homology on the amino acid level and is more closely related to nel than NRP1 having 50% homology. The cysteine residues in cysteine-rich domains and EGF-like repeats were found completely conserved.

25

[0280]

The most remarkable difference between the NRPs and the chick protein was that the human homologs lack the putative transmembrane domain of nel. However, even in this lacking region, the nucleotide sequences of NRPs
5 were very similar to that of nel. Furthermore, the two NRPs each possessed six EGF-like repeats, whereas nel has only five.

[0281]

Other unique motifs of nel as reported by
10 Matsuhashi et al. [Matsuhashi, S., et al., Dev. Dynamics, 203, 212-222 (1995)] were also found in the NRPs at equivalent positions. Since as mentioned above, it was shown that the two deduced NRP peptides are not
transmembrane proteins, the NRPs might be secretory
15 proteins or proteins anchored to membranes as a result of posttranslational modification.

[0282]

The present inventors speculate that NRPs might function as ligands by stimulating other molecules such
20 as EGF receptors. The present inventors further found that an extra EGF-like repeat could be encoded in nel upon frame shifting of the membrane domain region of nel.

[0283]

When paralleled and compared with NRP2 and nel,
25 the frame-shifted amino acid sequence showed similarities

over the whole range of NRP2 and of nel, suggesting that NRP2 might be a human counterpart of nel. In contrast, NRP1 is considered to be not a human counterpart of nel but a homologous gene.

5 [0284]

(2) Northern blot analysis

Northern blot analysis was carried out as described in Example 1-(2). Thus, the entire sequences of both clones cDNAs were amplified by PCR, the PCR
10 products were purified and labeled with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim) and human normal tissues were examined for NRP mRNA expression using an MTN blot with the labeled products as two probes.

15 [0285]

Sixteen adult tissues and four human fetal tissues were examined for the expression pattern of two NRPs.

[0286]

20 As a result of the Northern blot analysis, it was found that a 3.5 kb transcript of NRP1 was weakly expressed in fetal and adult brain and kidney. A 3.6 kb transcript of NRP2 was strongly expressed in adult and fetal brain alone, with weak expression thereof in fetal
25 kidney as well.

[0287]

This suggests that NRPs might play a brain-specific role, for example as signal molecules for growth regulation. In addition, these genes might have a
5 particular function in kidney.

[0288]

(3) Chromosomal mapping of NRP1 gene and NRP2 gene by FISH

Chromosomal mapping of the NRP1 gene and NRP2
10 gene was performed by FISH as described in Example 1-(3).

[0289]

As a result, it was revealed that the chromosomal locus of the NRP1 gene is localized to 11p15.1-p15.2 and the chromosomal locus of the NRP2 gene
15 to 12q13.11-q13.12.

[0290]

According to the present invention, the novel human NRP1 gene and NRP2 gene are provided and the use of said genes makes it possible to detect the expression of
20 said genes in various tissues and produce the human NRP1 and NRP2 proteins by the technology of genetic engineering. They can further be used in the study of the brain neurotransmission system, diagnosis of various diseases related to neurotransmission in the brain, and
25 the screening and evaluation of drugs for the treatment

and prevention of such diseases. Furthermore, the possibility is suggested that these EGF domain-containing NRPs act as growth factors in brain, hence they may be useful in the diagnosis and treatment of various kinds of intracerebral tumor and effective in nerve regeneration in cases of degenerative nervous diseases.

[0291]

[Example 11] GSPT1-related protein (GSPT1-TK) gene

(1) GSPT1-TK gene cloning and DNA sequencing

10 The human GSPT1 gene is one of the human homologous genes of the yeast GST1 gene that encodes the GTP-binding protein essential for the G1 to S phase transition in the cell cycle. The yeast GST1 gene, first identified as a protein capable of complementing a temperature-sensitive *gst1* (G1-to-S transition) mutant of Saccharomyces cerevisiae, was isolated from a yeast genomic library [Kikuchi, Y., Shimatake, H. and Kikuchi, A., EMBO J., 7, 1175-1182 (1988)] and encoded a protein with a target site of cAMP-dependent protein kinases and

15

20 a GTPase domain.

[0292]

The human GSPT1 gene was isolated from a KB cell cDNA library by hybridization using the yeast GST1 gene as a probe [Hoshino, S., Miyazawa, H., Enomoto, T., Hanaoka, F., Kikuchi, Y., Kikuchi, A. and Ui, M., EMBO

25

J., 8, 3807-3814 (1989)]. The deduced protein of said GSPT1 gene, like yeast GST1, has a GTP-binding domain and a GTPase activity center, and plays an important role in cell proliferation.

5 [0293]

Furthermore, a breakpoint for chromosome rearrangement has been observed in the GSPT1 gene located in the chromosomal locus 16p13.3 in patients with acute nonlymphocytic leukemia (ANLL) [Ozawa, K., Murakami, Y.,
10 Eki, T., Yokoyama, K. Soeda, E., Hoshino, S. Ui, M. and Hanaoka, F., Somatic Cell and Molecular Genet., 18, 189-194 (1992)].

[0294]

cDNA clones were randomly selected from a human
15 fetal brain cDNA library and subjected to sequence analysis as described in Example 1-(1) and database searching was performed and, as a result, a clone having a 0.3 kb cDNA sequence highly homologous to the above-mentioned GSPT1 gene was found and named GEN-077A09. The
20 GEN-077A09 clone seemed to be lacking in the 5' region, so that 5' RACE was carried out in the same manner as in Example 2-(2) to obtain the entire coding region.

[0295]

The primers used for the 5' RACE were P1 and P2
25 primers respectively having the nucleotide sequences

shown in Table 11 as designed based on the known cDNA sequence of the above-mentioned cDNA, and the anchor primer used was the one attached to the commercial kit.

35 cycles of PCR were performed under the following

5 conditions: 94°C for 45 seconds, 58°C for 45 seconds and 72°C for 2 minutes. Finally, elongation reaction was carried out at 72°C for 7 minutes.

[0296]

[Table 10]

10

Primer	Nucleotide sequence
P1 primer	5'-GATTTGTGCTCAATAATCACTATCTGAA-3'
P2 primer	5'-GGTTACTAGGATCACAAAGTATGAATTCTGGAA-3'

15

[0297]

Several of the 5' RACE clones obtained from the above PCR were sequenced and the base sequence of that
20 cDNA clone showing overlapping between the 5' RACE clones and the GEN-077A09 clone was determined to thereby reveal the sequence regarded as covering the entire coding region. This was named GSPT1-related protein "GSPT1-TK gene".

25

[0298]

The GSPT1-TK gene was found to contain an open reading frame of 1,497 nucleotides, as shown under SEQ ID NO:41. The amino acid sequence deduced therefrom

contained 499 amino acid residues, as shown under SEQ ID NO:40.

[0299]

The nucleotide sequence of the whole cDNA clone
5 of the GSPT1-TK gene was found to comprise 2,057
nucleotides, as shown under SEQ ID NO:42, and the
molecular weight was calculated at 55,740 daltons.

[0300]

The first methionine code (ATG) in the open
10 reading frame had no in-frame termination codon but this
ATG was surrounded by a sequence similar to the Kozak
consensus sequence for translational initiation.
Therefore, it was concluded that this ATG triplet
occurring in positions 144-146 of the relevant sequence
15 is the initiation codon.

[0301]

Furthermore, a polyadenylation signal, AATAAA,
was observed 13 nucleotides upstream from the
polyadenylation site.

20 [0302]

Human GSPT1-TK contains a glutamic acid rich
region near the N terminus, and 18 of 20 glutamic acid
residues occurring in this region of human GSPT1-TK are
conserved and align perfectly with those of the human
25 GSPT1 protein. Several regions (G1, G2, G3, G4 and G5)

of GTP-binding proteins that are responsible for guanine nucleotide binding and hydrolysis were found conserved in the GSPT1-TK protein just as in the human GSPT1 protein.

[0303]

5 Thus, the DNA sequence of human GSPT1-TK was found 89.4% identical, and the amino acid sequence deduced therefrom 92.4% identical, with the corresponding sequence of human GSPT1 which supposedly plays an important role in the G1 to S phase transition in the
10 cell cycle. Said amino acid sequence showed 50.8% identity with that of yeast GST1.

[0304]

(2) Northern blot analysis

Northern blot analysis was carried out as
15 described in Example 1-(2). Thus, the GEN-077A09 cDNA clone was amplified by PCR, the PCR product was purified and labeled with [³²P]-dCTP (random-primed DNA labeling kit, Boehringer Mannheim), and normal human tissues were examined for the expression of GSPT1-TK mRNA therein
20 using an MTN blot with the labeled product as a probe.

[0305]

As a result of the Northern blot analysis, a 2.7 kb major transcript was detected in various tissues. The level of human GSPT1-TK expression seemed highest in
25 brain and in testis.

[0306]

(3) Chromosome mapping of GSPT1-TK gene by FISH

Chromosome mapping of the GSPT1-TK gene was performed by FISH as described in Example 1-(3).

5 [0307]

As a result, it was found that the GSPT1-TK gene is localized at the chromosomal locus 19p13.3. In this chromosomal localization site, reciprocal location has been observed very frequently in cases of acute lymphocytic leukemia (ALL) and acute myeloid leukemia (AML). In addition, it is reported that acute non-lymphocytic leukemia (ANLL) is associated with re-arrangements involving the human GSPT1 region [Ozawa, K., Murakami, Y., Eki, T., Yokoyama, K., Soeda, E., Hoshino, S., Ui, M. and Hanaoka, F., Somatic Cell and Molecular Genet., 18, 189-194 (1992)].

10

15

[0308]

In view of the above, it is suggested that this gene is the best candidate gene associated with ALL and AML.

20

[0309]

In accordance with the present invention, the novel human GSPT1-TK gene is provided and the use of said gene makes it possible to detect the expression of said gene in various tissues and produce the human GSPT1-TK

25

protein by the technology of genetic engineering. These
can be used in the studies of cell proliferation, as
mentioned above, and further make it possible to diagnose
various diseases associated with the chromosomal locus of
5 this gene, for example acute myelocytic leukemia. This
is because translocation of this gene may result in
decomposition of the GSPT1-TK gene and further some or
other fused protein expressed upon said translocation may
cause such diseases.

10 [0310]

Furthermore, it is expected that diagnosis and
treatment of said diseases can be made possible by
producing antibodies to such fused protein, revealing the
intracellular localization of said protein and examining
15 its expression specific to said diseases. Therefore, it
is also expected that the use of the gene of the present
invention makes it possible to screen out and evaluate
drugs for the treatment and prevention of said diseases.

[0311]

[SEQUENCE LISTING]

[0312]

SEQ ID NO:1

SEQUENCE CHARACTERISTICS:

LENGTH: 122 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met	Glu	Leu	Gly	Glu	Asp	Gly	Ser	Val	Tyr	Lys	Ser	Ile	Leu	Val	Thr
1				5					10					15	
Ser	Gln	Asp	Lys	Ala	Pro	Ser	Val	Ile	Ser	Arg	Val	Leu	Lys	Lys	Asn
			20					25					30		
Asn	Arg	Asp	Ser	Ala	Val	Ala	Ser	Glu	Tyr	Glu	Leu	Val	Gln	Leu	Leu
		35					40					45			
Pro	Gly	Glu	Arg	Glu	Leu	Thr	Ile	Pro	Ala	Ser	Ala	Asn	Val	Phe	Tyr
	50					55					60				
Pro	Met	Asp	Gly	Ala	Ser	His	Asp	Phe	Leu	Leu	Arg	Gln	Arg	Arg	Arg
65					70					75					80
Ser	Ser	Thr	Ala	Thr	Pro	Gly	Val	Thr	Ser	Gly	Pro	Ser	Ala	Ser	Gly
				85					90					95	
Thr	Pro	Pro	Ser	Glu	Gly	Gly	Gly	Gly	Ser	Phe	Pro	Arg	Ile	Lys	Ala
			100					105						110	
Thr	Gly	Arg	Lys	Ile	Ala	Arg	Ala	Leu	Phe						
			115				120								

[0313]

SEQ ID NO:2

SEQUENCE CHARACTERISTICS:

LENGTH: 366 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (cDNA)

SEQUENCE DESCRIPTION:

ATGGAGTTGG GGAAGATGG CAGTGICTAT AAGAGCATTT TGGTGACAAG CCAGGACAAG	60
GCTCCAAGTG TCATCAGTCG TGICCTTAAG AAAACAATC GTGACTCTGC AGTGGCTTCA	120
GAGTATGAGC TGGTACAGCT GCTACCAGGG GAGCGAGAGC TGACTATCCC AGCTCGGCT	180
AATGTATTCT ACCCATGGA TGGAGCTTCA CACGATTTC TCCTGCGGCA GCGGCGAAGG	240
TCCTCTACTG CTACACCTGG CGTCACCAGT GGCCCGTCTG CCTCAGGAAC TCCTCCGAGT	300
GAGGGAGGAG GGGGCTCCTT TCCCAGGATC AAGGCCACAG GGAGGAAGAT TGCACGGGCA	360
CTGTTT	366

[0314]

SEQ ID NO:3

SEQUENCE CHARACTERISTICS:

LENGTH: 842 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-501D08

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 28..393

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

CCCACGAGCC GTATCATCCG AGTCCAG ATG GAG TTG GGG GAA GAT GGC AGT	51
Met Glu Leu Gly Glu Asp Gly Ser	
1 5	
GTC TAT AAG AGC ATT TTG GTG ACA AGC CAG GAC AAG GCT CCA AGT GTC	99
Val Tyr Lys Ser Ile Leu Val Thr Ser Gln Asp Lys Ala Pro Ser Val	
10 15 20	
ATC AGT CGT GTC CTT AAG AAA AAC AAT CGT GAC TCT GCA GTG GCT TCA	147
Ile Ser Arg Val Leu Lys Lys Asn Asn Arg Asp Ser Ala Val Ala Ser	
25 30 35 40	
GAG TAT GAG CTG GTA CAG CTG CTA CCA GGG GAG CGA GAG CTG ACT ATC	195
Glu Tyr Glu Leu Val Gln Leu Leu Pro Gly Glu Arg Glu Leu Thr Ile	
45 50 55	
CCA GCC TCG GCT AAT GTA TTC TAC CCC ATG GAT GGA GCT TCA CAC GAT	243
Pro Ala Ser Ala Asn Val Phe Tyr Pro Met Asp Gly Ala Ser His Asp	
60 65 70	
TTC CTC CTG CGG CAG CGG CGA AGG TCC TCT ACT GCT ACA CCT GGC GTC	291
Phe Leu Leu Arg Gln Arg Arg Arg Ser Ser Thr Ala Thr Pro Gly Val	
75 80 85	
ACC AGT GGC CCG TCT GCC TCA GGA ACT CCT CCG AGT GAG GGA GGA GGG	339
Thr Ser Gly Pro Ser Ala Ser Gly Thr Pro Pro Ser Glu Gly Gly Gly	
90 95 100	
GGC TCC TTT CCC AGG ATC AAG GCC ACA GGG AGG AAG ATT GCA CGG GCA	387
Gly Ser Phe Pro Arg Ile Lys Ala Thr Gly Arg Lys Ile Ala Arg Ala	
105 110 115 120	
CTG TTC TGAGGAGGAA GCCCCTTTTT TTACAGAAGT CATGGTGTTT ATACCAGATG	443
Leu Phe	
TGGGTAGCCA TCCTGAATGG TGGCAATTAT ATCACATTGA GACAGAAATT CAGAAAGGGA	503

```

GCCAGCCACC CTGGGGCAGT GAAGTGCCAC TGGTTTACCA GACAGCTGAG AAATCCAGCC      563
CTGTCGGAAC TGGTGTCTTA TAACCAAGTT GGATACCTGT GTATAGCTTG CCACCTTCCA      623
TGAGTGCAGC ACACAGGTAG TGCTGGAAAA ACGCATCAGT TTCTGATTCT TGGCCATATC      683
CTAACATGCA AGGGCCAAGC AAAGGCTTCA AGGCTCTGAG CCCCAGGGCA GAGGGGAATG      743
GCAAAATGTA GGTCTTGGCA GGAGCTCTTC TTCCCACTCT GGGGGTTTCT ATCACTGTGA      803
CAACACTAAG ATAATAAACC AAAACACTAC CTGAATTCT                                842

```

[0315]

SEQ ID NO:4

SEQUENCE CHARACTERISTICS:

LENGTH: 193 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

```

Met Glu Leu Glu Leu Tyr Gly Val Asp Asp Lys Phe Tyr Ser Lys Leu
 1              5              10              15
Asp Gln Glu Asp Ala Leu Leu Gly Ser Tyr Pro Val Asp Asp Gly Cys
      20              25              30
Arg Ile His Val Ile Asp His Ser Gly Ala Arg Leu Gly Glu Tyr Glu
      35              40              45
Asp Val Ser Arg Val Glu Lys Tyr Thr Ile Ser Gln Glu Ala Tyr Asp
      50              55              60
Gln Arg Gln Asp Thr Val Arg Ser Phe Leu Lys Arg Ser Lys Leu Gly
      65              70              75              80
Arg Tyr Asn Glu Glu Glu Arg Ala Gln Gln Glu Ala Glu Ala Ala Gln
      85              90              95
Arg Leu Ala Glu Glu Lys Ala Gln Ala Ser Ser Ile Pro Val Gly Ser
      100              105              110

```


-109-

Arg Cys Glu Val Arg Ala Ala Gly Gln Ser Pro Arg Arg Gly Thr Val
115 120 125
Met Tyr Val Gly Leu Thr Asp Phe Lys Pro Gly Tyr Trp Ile Gly Val
130 135 140
Arg Tyr Asp Glu Pro Leu Gly Lys Asn Asp Gly Ser Val Asn Gly Lys
145 150 155 160
Arg Tyr Phe Glu Cys Gln Ala Lys Tyr Gly Ala Phe Val Lys Pro Ala
165 170 175
Val Val Thr Val Gly Asp Phe Pro Glu Glu Asp Tyr Gly Leu Asp Glu
180 185 190
Ile

[0316]

SEQ ID NO:5

SEQUENCE CHARACTERISTICS:

LENGTH: 579 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (cDNA)

SEQUENCE DESCRIPTION:

ATGGA	ACTGG	AGCTG	TATGG	AGTTG	ACGAC	AAGTT	TCTACA	GCAAG	CTGGA	TCAAG	AGGAT	60
GCGCT	CCTGG	GCTCCT	ACCC	TGTAG	ATGAC	GGCTG	CCGCA	TCCAC	GTCAT	TGACC	CACAGT	120
GGCGC	CCGCC	TTGGT	GAGTA	TGAGG	ACGTG	TCCCG	GGTGG	AGAAG	TACAC	GATCT	CACAA	180
GAAGC	CTACG	ACCAG	AGGCA	AGACAC	GGTC	CGCTC	TTTCC	TGAAG	CGCAG	CAAGC	TCGGC	240
CGGTAC	AACG	AGGAG	GAGCG	GGCTC	AGCAG	GAGGC	CGAGG	CCGCC	CAGCG	CCTGG	CCGAG	300
GAGAAG	GCCC	AGGCC	AGCTC	CATCCC	CGTG	GCCAG	CCGCT	GTGAG	GTGCG	GGCGG	CGGGA	360
CAATCCC	CTC	GCCGG	GGCAC	CGTCAT	GTAT	GTAGG	TCTCA	CAGATT	TCAA	GCCTG	GCTAC	420

-110-

TGGATTGGTG TCCGCTATGA TGAGCCACTG GGGAAAAATG ATGGCAGTGT GAATGGGAAA	480
CGCTACTTCG AATGCCAGGC CAAGTATGGC GCCTTTGTCA AGCCAGCAGT CGTGACGGTG	540
GGGGACTTCC CGGAGGAGGA CTACGGGTG GACGAGATA	579

[0317]

SEQ ID NO:6

SEQUENCE CHARACTERISTICS:

LENGTH: 1015 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-080G01

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 274..852

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

TGATTGGTCA GGCACGGAGC AGGAGGCGGG CTGATAGCCC AGCAGCAGCA GCGGCGGCGG	60
CGGCTGCGGA GCGGGTGTGA GGCGGCTGGA CCGCGCTGCA GGCATCCGCG GGCGCGGCAA	120
GATGGAGGTG ACGGGGGTGT CGGCACCACG GTGACCGTTT TCATCAGCAG CTCCCTCAGC	180
ACCTTCCGCT CCGAGAAGCG ATACAGCCGC AGCCTCACCA TCGCTGAGTT CAAGTGTAAG	240
CTGGAGTTGC TGGTGGGCAG CCCTGCTTCC TGC ATG GAA CTG GAG CTG TAT GGA	294
Met Glu Leu Glu Leu Tyr Gly	
1 5	

GTT GAC GAC AAG TTC TAC AGC AAG CTG GAT CAA GAG GAT GCG CTC CTG Val Asp Asp Lys Phe Tyr Ser Lys Leu Asp Gln Glu Asp Ala Leu Leu 10 15 20	342
GGC TCC TAC CCT GTA GAT GAC GGC TGC CGC ATC CAC GTC ATT GAC CAC Gly Ser Tyr Pro Val Asp Asp Gly Cys Arg Ile His Val Ile Asp His 25 30 35	390
AGT GGC GCC CGC CTT GGT GAG TAT GAG GAC GTG TCC CGG GTG GAG AAG Ser Gly Ala Arg Leu Gly Glu Tyr Glu Asp Val Ser Arg Val Glu Lys 40 45 50 55	438
TAC ACG ATC TCA CAA GAA GCC TAC GAC CAG AGG CAA GAC ACG GTC CGC Tyr Thr Ile Ser Gln Glu Ala Tyr Asp Gln Arg Gln Asp Thr Val Arg 60 65 70	486
TCT TTC CTG AAG CGC AGC AAG CTC GGC CGG TAC AAC GAG GAG GAG CGG Ser Phe Leu Lys Arg Ser Lys Leu Gly Arg Tyr Asn Glu Glu Glu Arg 75 80 85	534
GCT CAG CAG GAG GCC GAG GCC GCC CAG CGC CTG GCC GAG GAG AAG GCC Ala Gln Gln Glu Ala Glu Ala Ala Gln Arg Leu Ala Glu Glu Lys Ala 90 95 100	582
CAG GCC AGC TCC ATC CCC GTG GGC AGC CGC TGT GAG GTG CGG GCG GCG Gln Ala Ser Ser Ile Pro Val Gly Ser Arg Cys Glu Val Arg Ala Ala 105 110 115	630
GGA CAA TCC CCT CGC CGG GGC ACC GTC ATG TAT GTA GGT CTC ACA GAT Gly Gln Ser Pro Arg Arg Gly Thr Val Met Tyr Val Gly Leu Thr Asp 120 125 130 135	678
TTC AAG CCT GGC TAC TGG ATT GGT GTC CGC TAT GAT GAG CCA CTG GGG Phe Lys Pro Gly Tyr Trp Ile Gly Val Arg Tyr Asp Glu Pro Leu Gly 140 145 150	726
AAA AAT GAT GGC AGT GTG AAT GGG AAA CGC TAC TTC GAA TGC CAG GCC Lys Asn Asp Gly Ser Val Asn Gly Lys Arg Tyr Phe Glu Cys Gln Ala 155 160 165	774
AAG TAT GGC GCC TTT GTC AAG CCA GCA GTC GTG ACG GTG GGG GAC TTC Lys Tyr Gly Ala Phe Val Lys Pro Ala Val Val Thr Val Gly Asp Phe 170 175 180	822
CCG GAG GAG GAC TAC GGG TTG GAC GAG ATA TGACACCTAA GGAATTCCTCC Pro Glu Glu Asp Tyr Gly Leu Asp Glu Ile 185 190	872
TGCTTCAGCT CCTAGCTCAG CCACTGACTG CCCCTCCTGT GTGTGCCCCAT GGCCCTTTTC	932

-112-

TCCTGACCCC ATTTTAATTT TATTCATTTT TTCCTTTGCC ATTGATTTT GAGACTCATG 992
CATTAAATTC ACTAGAAACC CAG 1015

[0318]

SEQ ID NO:7

SEQUENCE CHARACTERISTICS:

LENGTH: 128 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met	Thr	Glu	Ala	Asp	Val	Asn	Pro	Lys	Ala	Tyr	Pro	Leu	Ala	Asp	Ala
1				5					10					15	
His	Leu	Thr	Lys	Lys	Leu	Leu	Asp	Leu	Val	Gln	Gln	Ser	Cys	Asn	Tyr
			20					25					30		
Lys	Gln	Leu	Arg	Lys	Gly	Ala	Asn	Glu	Ala	Thr	Lys	Thr	Leu	Asn	Arg
		35					40					45			
Gly	Ile	Ser	Glu	Phe	Ile	Val	Met	Ala	Ala	Asp	Ala	Glu	Pro	Leu	Glu
	50					55				60					
Ile	Ile	Leu	His	Leu	Pro	Leu	Leu	Cys	Glu	Asp	Lys	Asn	Val	Pro	Tyr
65					70					75					80
Val	Phe	Val	Arg	Ser	Lys	Gln	Ala	Leu	Gly	Arg	Ala	Cys	Gly	Val	Ser
				85					90					95	
Arg	Pro	Val	Ile	Ala	Cys	Ser	Val	Thr	Ile	Lys	Glu	Gly	Ser	Gln	Leu
			100					105					110		
Lys	Gln	Gln	Ile	Gln	Ser	Ile	Gln	Gln	Ser	Ile	Glu	Arg	Leu	Leu	Val
		115					120					125			

[0319]

SEQ ID NO:8

SEQUENCE CHARACTERISTICS:

LENGTH: 384 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGACTGAGG CTGATGTGAA TCCAAAGGCC TATCCCCCTG CCGATGCCCA CCTCACCAAG	60
AAGCTACTGG ACCTCGTTCA GCAGTCATGT AACTATAAGC AGCTTCGGAA AGGAGCCAAT	120
GAGGCCACCA AAACCCTCAA CAGGGGCATC TCTGAGTICA TCGTGATGGC TGCAGACGCC	180
GAGCCACTGG AGATCATTCT GCACCTGCCG CTGCTGTGTG AAGACAAGAA TGTGCCCTAC	240
GTGTTTGTGC GCTCCAAGCA GCCCCTGGGG AGAGCCTGTG GGGTCTCCAG GCCTGTCATC	300
GCCTGTTCTG TCACCATCAA AGAAGGCTCG CAGCTGAAAC AGCAGATCCA ATCCATTCAG	360
CAGTCCATTG AAAGGCTCTT AGTC	384

[0320]

SEQ ID NO:9

SEQUENCE CHARACTERISTICS:

LENGTH: 1493 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-025F07

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 95..478

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

ATCCGIGTCC TTGCGGTGCT GGGCAGCAGA CCGTCCAAAC CGACACGCGT GGTATCCTCG	60
CGGTGTCCGG CAAGAGACTA CCAAGACAGA CGCT ATG ACT GAG GCT GAT GTG	112
Met Thr Glu Ala Asp Val	
1 5	
AAT CCA AAG GCC TAT CCC CTT GCC GAT GCC CAC CTC ACC AAG AAG CTA	160
Asn Pro Lys Ala Tyr Pro Leu Ala Asp Ala His Leu Thr Lys Lys Leu	
10 15 20	
CTG GAC CTC GTT CAG CAG TCA TGT AAC TAT AAG CAG CTT CGG AAA GGA	208
Leu Asp Leu Val Gln Gln Ser Cys Asn Tyr Lys Gln Leu Arg Lys Gly	
25 30 35	
GCC AAT GAG GCC ACC AAA ACC CTC AAC AGG GGC ATC TCT GAG TTC ATC	256
Ala Asn Glu Ala Thr Lys Thr Leu Asn Arg Gly Ile Ser Glu Phe Ile	
40 45 50	
GTG ATG GCT GCA GAC GCC GAG CCA CTG GAG ATC ATT CTG CAC CTG CCG	304
Val Met Ala Ala Asp Ala Glu Pro Leu Glu Ile Ile Leu His Leu Pro	
55 60 65 70	
CTG CTG TGT GAA GAC AAG AAT GTG CCC TAC GTG TTT GTG CGC TCC AAG	352
Leu Leu Cys Glu Asp Lys Asn Val Pro Tyr Val Phe Val Arg Ser Lys	
75 80 85	
CAG GCC CTG GGG AGA GCC TGT GGG GTC TCC AGG CCT GTC ATC GCC TGT	400
Gln Ala Leu Gly Arg Ala Cys Gly Val Ser Arg Pro Val Ile Ala Cys	
90 95 100	
TCT GTC ACC ATC AAA GAA GGC TCG CAG CTG AAA CAG CAG ATC CAA TCC	448
Ser Val Thr Ile Lys Glu Gly Ser Gln Leu Lys Gln Gln Ile Gln Ser	
105 110 115	
ATT CAG CAG TCC ATT GAA AGG CTC TTA GTC TAAACCTGTG GCCTCTGCCA	498
Ile Gln Gln Ser Ile Glu Arg Leu Leu Val	
120 125	

CGTGCTCCCT GCCAGCTTCC CCCCTGAGGT TGTGTATCAT ATTATCTGTG TTAGCATGTA	558
GTATTTTCAG CTACTCTCTA TTGTTATAAA ATGTAGTACT AAATCTGGTT TCTGGATTTT	618
TGTGTTGTTT TTGTTCTGTT TTACAGGGTT GCTATCCCCC TTCCTTTCCT CCCTCCCTCT	678
GCCATCCTTC ATCCTTTTAT CCTCCCTTTT TGAACAAGT GTTCAGAGCA GACAGAAGCA	738
GGGTGGTGGC ACCGTTGAAA GGCAGAAAGA GCCAGGAGAA AGCTGATGGA GCCAGGACAG	798
AGATCTGGTT CCAGCTTTCA GCCACTAGCT TCCTGTTGTG TCGGGGGTGT GGTGGAATTA	858
AACAGCATTC ATTGTGTGTC CCTGTGCCTG GCACACAGAA TCATTATAC GTGTTCAAGT	918
GATCAAGGGG TTTCAATTTG TCTTGGGGGA TTAGGTATCA TTTGGGGAGG AAGCATGTGT	978
TCTGTGAGGT TGTTGCGCTA TGTCCAAGTG TCGTTTACTA ATGTACCCCT GCTGTTTGCT	1038
TTTGGTAATG TGATGTTGAT GTTCTCCCCC TACCCACAAC CATGCCCTTG AGGGTAGCAG	1098
GGCAGCAGCA TACCAAAGAG ATGTGCTGCA GGACTCCGGA GGCAGCCTGG GTGGGTGAGC	1158
CATGGGGCAG TTGACCTGGG TCTTGAAAGA GTCGGGAGTG ACAAGCTCAG AGAGCATGAA	1218
CTGATGCTGG CATGAAGGAT TCCAGGAAGA TCATGGAGAC CTGGCTGGTA GCTGTAACAG	1278
AGATGGTGGA GTCCAAGGAA ACAGCCTGTC TCTGGTGAAT GGGACTTTCT TTGGTGGACA	1338
CTTGGCACCA GCTCTGAGAG CCTTCCCCCT GTGTCTGCC ACCATGTGGG TCAGATGTAC	1398
TCTCTGTAC ATGAGGAGAG TGCTAGTTCA TGIGTTCTCC ATTCTTGIGA GCATCCTAAT	1458
AAATCTGTTT CATTTTGAAA AAAAAAAAAA AAAAA	1493

[0321]

SEQ ID NO:10

SEQUENCE CHARACTERISTICS:

LENGTH: 711 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met	Pro	Ala	Asp	Val	Asn	Leu	Ser	Gln	Lys	Pro	Gln	Val	Leu	Gly	Pro	1	5	10	15
Glu	Lys	Gln	Asp	Gly	Ser	Cys	Glu	Ala	Ser	Val	Ser	Phe	Glu	Asp	Val	20	25	30	
Thr	Val	Asp	Phe	Ser	Arg	Glu	Glu	Trp	Gln	Gln	Leu	Asp	Pro	Ala	Gln	35	40	45	
Arg	Cys	Leu	Tyr	Arg	Asp	Val	Met	Leu	Glu	Leu	Tyr	Ser	His	Leu	Phe	50	55	60	
Ala	Val	Gly	Tyr	His	Ile	Pro	Asn	Pro	Glu	Val	Ile	Phe	Arg	Met	Leu	65	70	75	80
Lys	Glu	Lys	Glu	Pro	Arg	Val	Glu	Glu	Ala	Glu	Val	Ser	His	Gln	Arg	85	90	95	
Cys	Gln	Glu	Arg	Glu	Phe	Gly	Leu	Glu	Ile	Pro	Gln	Lys	Glu	Ile	Ser	100	105	110	
Lys	Lys	Ala	Ser	Phe	Gln	Lys	Asp	Met	Val	Gly	Glu	Phe	Thr	Arg	Asp	115	120	125	
Gly	Ser	Trp	Cys	Ser	Ile	Leu	Glu	Glu	Leu	Arg	Leu	Asp	Ala	Asp	Arg	130	135	140	
Thr	Lys	Lys	Asp	Glu	Gln	Asn	Gln	Ile	Gln	Pro	Met	Ser	His	Ser	Ala	145	150	155	160
Phe	Phe	Asn	Lys	Lys	Thr	Leu	Asn	Thr	Glu	Ser	Asn	Cys	Glu	Tyr	Lys	165	170	175	
Asp	Pro	Gly	Lys	Met	Ile	Arg	Thr	Arg	Pro	His	Leu	Ala	Ser	Ser	Gln	180	185	190	
Lys	Gln	Pro	Gln	Lys	Cys	Cys	Leu	Phe	Thr	Glu	Ser	Leu	Lys	Leu	Asn	195	200	205	
Leu	Glu	Val	Asn	Gly	Gln	Asn	Glu	Ser	Asn	Asp	Thr	Glu	Gln	Leu	Asp	210	215	220	
Asp	Val	Val	Gly	Ser	Gly	Gln	Leu	Phe	Ser	His	Ser	Ser	Ser	Asp	Ala	225	230	235	240
Cys	Ser	Lys	Asn	Ile	His	Thr	Gly	Glu	Thr	Phe	Cys	Lys	Gly	Asn	Gln	245	250	255	

Cys Arg Lys Val Cys Gly His Lys Gln Ser Leu Lys Gln His Gln Ile
 260 265 270
 His Thr Gln Lys Lys Pro Asp Gly Cys Ser Glu Cys Gly Gly Ser Phe
 275 280 285
 Thr Gln Lys Ser His Leu Phe Ala Gln Gln Arg Ile His Ser Val Gly
 290 295 300
 Asn Leu His Glu Cys Gly Lys Cys Gly Lys Ala Phe Met Pro Gln Leu
 305 310 315 320
 Lys Leu Ser Val Tyr Leu Thr Asp His Thr Gly Asp Ile Pro Cys Ile
 325 330 335
 Cys Lys Glu Cys Gly Lys Val Phe Ile Gln Arg Ser Glu Leu Leu Thr
 340 345 350
 His Gln Lys Thr His Thr Arg Lys Lys Pro Tyr Lys Cys His Asp Cys
 355 360 365
 Gly Lys Ala Phe Phe Gln Met Leu Ser Leu Phe Arg His Gln Arg Thr
 370 375 380
 His Ser Arg Glu Lys Leu Tyr Glu Cys Ser Glu Cys Gly Lys Gly Phe
 385 390 395 400
 Ser Gln Asn Ser Thr Leu Ile Ile His Gln Lys Ile His Thr Gly Glu
 405 410 415
 Arg Gln Tyr Ala Cys Ser Glu Cys Gly Lys Ala Phe Thr Gln Lys Ser
 420 425 430
 Thr Leu Ser Leu His Gln Arg Ile His Ser Gly Gln Lys Ser Tyr Val
 435 440 445
 Cys Ile Glu Cys Gly Gln Ala Phe Ile Gln Lys Ala His Leu Ile Val
 450 455 460
 His Gln Arg Ser His Thr Gly Glu Lys Pro Tyr Gln Cys His Asn Cys
 465 470 475 480
 Gly Lys Ser Phe Ile Ser Lys Ser Gln Leu Asp Ile His His Arg Ile
 485 490 495
 His Thr Gly Glu Lys Pro Tyr Glu Cys Ser Asp Cys Gly Lys Thr Phe
 500 505 510
 Thr Gln Lys Ser His Leu Asn Ile His Gln Lys Ile His Thr Gly Glu

515		520		525
Arg His His Val Cys Ser Glu Cys Gly Lys Ala Phe Asn Gln Lys Ser				
530		535		540
Ile Leu Ser Met His Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Lys				
545		550		555
Cys Ser Glu Cys Gly Lys Ala Phe Thr Ser Lys Ser Gln Phe Lys Glu				
		565		570
His Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Val Cys Thr Glu Cys				
		580		585
Gly Lys Ala Phe Asn Gly Arg Ser Asn Phe His Lys His Gln Ile Thr				
		595		600
His Thr Arg Glu Arg Pro Phe Val Cys Tyr Lys Cys Gly Lys Ala Phe				
		610		615
Val Gln Lys Ser Glu Leu Ile Thr His Gln Arg Thr His Met Gly Glu				
		625		630
Lys Pro Tyr Glu Cys Leu Asp Cys Gly Lys Ser Phe Ser Lys Lys Pro				
		645		650
Gln Leu Lys Val His Gln Arg Ile His Thr Gly Glu Arg Pro Tyr Val				
		660		665
Cys Ser Glu Cys Gly Lys Ala Phe Asn Asn Arg Ser Asn Phe Asn Lys				
		675		680
His Gln Thr Thr His Thr Arg Asp Lys Ser Tyr Lys Cys Ser Tyr Ser				
		690		695
Val Lys Gly Phe Thr Lys Gln				
705		710		

[0322]

SEQ ID NO:11

SEQUENCE CHARACTERISTICS:

LENGTH: 2133 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGCCTGCTG ATGTGAATTT ATCCCAGAAG CCTCAGGTCC TGGGTCCAGA GAAGCAGGAT	60
GGATCTTGCG AGGCATCAGT GTCATTTGAG GACGTGACCG TGGACTTCAG CAGGGAGGAG	120
TGGCAGCAAC TGGACCCTGC CCAGAGATGC CTGTACCGGG ATGTGATGCT GGAGCTCTAT	180
AGCCATCTCT TCGCAGTGGG GTATCACATT CCCAACCCAG AGGTCATCTT CAGAATGCTA	240
AAAGAAAAGG AGCCGCGTGT GGAGGAGGCT GAAGTCTCAC ATCAGAGGTG TCAAGAAAGG	300
GAGTTTGGGC TTGAAATCCC ACAAAAGGAG ATTTCTAAGA AAGCTTCATT TCAAAAGGAT	360
ATGGTAGGTG AGTTCACAAG AGATGGTTCA TGGTGTTCCTA TTPTAGAAGA ACTGAGGCTG	420
GATGCTGACC GCACAAAGAA AGATGAGCAA AATCAAATTC AACCCATGAG TCACAGTGCT	480
TTCTTCAACA AGAAAACATT GAACACAGAA AGCAATTGTG AATATAAGGA CCCTGGGAAA	540
ATGATTCGCA CGAGGCCCCA CCTTGCTTCT TCACAGAAAC AACCTCAGAA ATGTTGCTTA	600
TTTACAGAAA GTTTGAAGCT GAACCTAGAA GTGAACGGTC AGAATGAAAG CAATGACACA	660
GAACAGCTTG ATGACGTTGT TGGGTCTGGT CAGCTATTCA GCCATAGCTC TTCTGATGCC	720
TGCAGCAAGA ATATTCATAC AGGAGAGACA TTTTGCAAAG GTAACCAAGT TAGAAAAGTC	780
TGTGGCCATA AACAGTCACT CAAGCAACAT CAAATTCATA CTCAGAAGAA ACCAGATGGA	840
TGTTCTGAAT GTGGGGGGAG CTTACCCAG AAGTCACACC TCTTTGCCCA ACAGAGAATT	900
CATAGTGTAG GAAACCTCCA TGAATGTGGC AAATGTGGAA AAGCCTTCAT GCCACAATA	960
AAACTCAGTG TATATCTGAC AGATCATACA GGTGATATAC CCTGTATATG CAAGGAATGT	1020
GGGAAGGTCT TTATTCAGAG ATCAGAATTG CTTACGCACC AGAAAACACA CACTAGAAAG	1080
AAGCCCTATA AATGCCATGA CTGTGGAAAA GCCTTTTTC CAGATGTTATC TCTCTTCAGA	1140
CATCAGAGAA CTCACAGTAG AGAAAACTC TATGAATGCA GTGAATGTGG CAAAGGCTTC	1200
TCCCAAACT CAACCCTCAT TATACATCAG AAAATTCATA CTGGTGAGAG ACAGTATGCA	1260

TGCAGTGAAT GTGGGAAAGC CTTTACCCAG AAGTCAACAC TCAGCTTGCA CCAGAGAATC	1320
CACTCAGGGC AGAAGTCCTA TGTGTGTATC GAATGCGGGC AGGCCTTCAT CCAGAAGGCA	1380
CACCTGATTG TCCATCAAAG AAGCCACACA GGAGAAAAAC CTTATCAGTG CCACAACGTG	1440
GGGAAATCCT TCATTTCCTA GTACAGCTT GATATACATC ATCGAATTCA TACAGGGGAG	1500
AAACCTTATG AATGCAGTGA CTGTGGAAAA ACCTTCACCC AAAAGTCACA CCTGAATATA	1560
CACCAGAAAA TTCATACTGG AGAAAGACAC CATGTATGCA GTGAATGCGG GAAAGCCTTC	1620
AACCAGAAGT CAATACTCAG CATGCATCAG AGAATTCACA CCGGAGAGAA GCCTTACAAA	1680
TGCAGTGAAT GTGGGAAAGC CTTCACTTCT AAGTCTCAAT TCAAAGAGCA TCAGCGAATT	1740
CACACGGGTG AGAAACCCTA TGTGTGCACT GAATGTGGGA AGGCCTTCAA CGGCAGGTCA	1800
AATTTCCATA AACATCAAAT AACTCACACT AGAGAGAGGC CTTTTGTCTG TTACAAATGT	1860
GGGAAGGCTT TTGTCCAGAA ATCAGAGTTG ATTACCCATC AAAGAACTCA CATGGGAGAG	1920
AAACCTATG AATGCCTTGA CTGTGGGAAA TCGTTCAGTA AGAAACCACA ACTCAAGGTG	1980
CATCAGCGAA TTCACACGGG AGAAAGACCT TATGTGTGTT CTGAATGTGG AAAGGCCTTC	2040
AACAACAGGT CAAACTTCAA TAAACACCAA ACAACTCATA CCAGAGACAA ATCTTACAAA	2100
TGCAGTTATT CTGTGAAAGG CTTTACCAAG CAA	2133

[0323]

SEQ ID NO:12

SEQUENCE CHARACTERISTICS:

LENGTH: 3754 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-076C09

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 346..2478

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

GCTAAGCCTA TGTGCTTAC TGGACGCTGA AGTGATTGGG AATATTAGCA GTGGGGGTTC	60
TGTAGGGTCA GGAAGGGGCG GCTGGCTTTG GGGGAGTGAT GAGGGGCTTG TTGGGGGTGG	120
GGGTGCGTGA TAAAGGGATT TCTCGGCTGA AGACGAGGCT GTGAGGCTTC TGCAGAACCC	180
CCAGGTCAGG CCACATCATT GAGGCTGCAG GATCTCTCTT CATAGCCCAG TACGACTCTC	240
CGCCGIGTCC CTGGTTGGAA AATCCAAACA CCTATCCAGC TTCTGGCTCC TGGGAAAAGT	300
GGAGTTGTCA GCAAGAGAGA CCGAGAGTAG AAGCCCAGAG TGGAG ATG CCT GCT	354
Met Pro Ala	
1	
GAT GTG AAT TTA TCC CAG AAG CCT CAG GTC CTG GGT CCA GAG AAG CAG	402
Asp Val Asn Leu Ser Gln Lys Pro Gln Val Leu Gly Pro Glu Lys Gln	
5 10 15	
GAT GGA TCT TGC GAG GCA TCA GTG TCA TTT GAG GAC GTG ACC GTG GAC	450
Asp Gly Ser Cys Glu Ala Ser Val Ser Phe Glu Asp Val Thr Val Asp	
20 25 30 35	
TTC AGC AGG GAG GAG TGG CAG CAA CTG GAC CCT GCC CAG AGA TGC CTG	498
Phe Ser Arg Glu Glu Trp Gln Gln Leu Asp Pro Ala Gln Arg Cys Leu	
40 45 50	
TAC CGG GAT GTG ATG CTG GAG CTC TAT AGC CAT CTC TTC GCA GTG GGG	546
Tyr Arg Asp Val Met Leu Glu Leu Tyr Ser His Leu Phe Ala Val Gly	
55 60 65	
TAT CAC ATT CCC AAC CCA GAG GTC ATC TTC AGA ATG CTA AAA GAA AAG	594
Tyr His Ile Pro Asn Pro Glu Val Ile Phe Arg Met Leu Lys Glu Lys	
70 75 80	
GAG CCG CGT GTG GAG GAG GCT GAA GTC TCA CAT CAG AGG TGT CAA GAA	642

Glu	Pro	Arg	Val	Glu	Glu	Ala	Glu	Val	Ser	His	Gln	Arg	Cys	Gln	Glu		
85						90					95						
AGG	GAG	TTT	GGG	CTT	GAA	ATC	CCA	CAA	AAG	GAG	ATT	TCT	AAG	AAA	GCT	690	
Arg	Glu	Phe	Gly	Leu	Glu	Ile	Pro	Gln	Lys	Glu	Ile	Ser	Lys	Lys	Ala		
100					105					110					115		
TCA	TTT	CAA	AAG	GAT	ATG	GTA	GGT	GAG	TTC	ACA	AGA	GAT	GGT	TCA	TGG	738	
Ser	Phe	Gln	Lys	Asp	Met	Val	Gly	Glu	Phe	Thr	Arg	Asp	Gly	Ser	Trp		
				120					125					130			
TGT	TCC	ATT	TTA	GAA	GAA	CTG	AGG	CTG	GAT	GCT	GAC	CGC	ACA	AAG	AAA	786	
Cys	Ser	Ile	Leu	Glu	Glu	Leu	Arg	Leu	Asp	Ala	Asp	Arg	Thr	Lys	Lys		
			135					140					145				
GAT	GAG	CAA	AAT	CAA	ATT	CAA	CCC	ATG	AGT	CAC	AGT	GCT	TTC	TTC	AAC	834	
Asp	Glu	Gln	Asn	Gln	Ile	Gln	Pro	Met	Ser	His	Ser	Ala	Phe	Phe	Asn		
	150						155					160					
AAG	AAA	ACA	TTG	AAC	ACA	GAA	AGC	AAT	TGT	GAA	TAT	AAG	GAC	CCT	GGG	882	
Lys	Lys	Thr	Leu	Asn	Thr	Glu	Ser	Asn	Cys	Glu	Tyr	Lys	Asp	Pro	Gly		
	165					170					175						
AAA	ATG	ATT	CGC	ACG	AGG	CCC	CAC	CTT	GCT	TCT	TCA	CAG	AAA	CAA	CCT	930	
Lys	Met	Ile	Arg	Thr	Arg	Pro	His	Leu	Ala	Ser	Ser	Gln	Lys	Gln	Pro		
180					185					190					195		
CAG	AAA	TGT	TGC	TTA	TTT	ACA	GAA	AGT	TTG	AAG	CTG	AAC	CTA	GAA	GTG	978	
Gln	Lys	Cys	Cys	Leu	Phe	Thr	Glu	Ser	Leu	Lys	Leu	Asn	Leu	Glu	Val		
				200					205					210			
AAC	GGT	CAG	AAT	GAA	AGC	AAT	GAC	ACA	GAA	CAG	CTT	GAT	GAC	GTT	GTT	1026	
Asn	Gly	Gln	Asn	Glu	Ser	Asn	Asp	Thr	Glu	Gln	Leu	Asp	Asp	Val	Val		
			215					220					225				
GGG	TCT	GGT	CAG	CTA	TTC	AGC	CAT	AGC	TCT	TCT	GAT	GCC	TGC	AGC	AAG	1074	
Gly	Ser	Gly	Gln	Leu	Phe	Ser	His	Ser	Ser	Ser	Asp	Ala	Cys	Ser	Lys		
		230					235					240					
AAT	ATT	CAT	ACA	GGA	GAG	ACA	TTT	TGC	AAA	GGT	AAC	CAG	TGT	AGA	AAA	1122	
Asn	Ile	His	Thr	Gly	Glu	Thr	Phe	Cys	Lys	Gly	Asn	Gln	Cys	Arg	Lys		
	245					250					255						
GTC	TGT	GGC	CAT	AAA	CAG	TCA	CTC	AAG	CAA	CAT	CAA	ATT	CAT	ACT	CAG	1170	
Val	Cys	Gly	His	Lys	Gln	Ser	Leu	Lys	Gln	His	Gln	Ile	His	Thr	Gln		
260					265					270					275		
AAG	AAA	CCA	GAT	GGA	TGT	TCT	GAA	TGT	GGG	GGG	AGC	TTC	ACC	CAG	AAG	1218	
Lys	Lys	Pro	Asp	Gly	Cys	Ser	Glu	Cys	Gly	Gly	Ser	Phe	Thr	Gln	Lys		

280	285	290	
TCA CAC CTC TTT GCC CAA CAG AGA ATT CAT AGT GTA GGA AAC CTC CAT Ser His Leu Phe Ala Gln Gln Arg Ile His Ser Val Gly Asn Leu His 295 300 305			1266
GAA TGT GGC AAA TGT GGA AAA GCC TTC ATG CCA CAA CTA AAA CTC AGT Glu Cys Gly Lys Cys Gly Lys Ala Phe Met Pro Gln Leu Lys Leu Ser 310 315 320			1314
GTA TAT CTG ACA GAT CAT ACA GGT GAT ATA CCC TGT ATA TGC AAG GAA Val Tyr Leu Thr Asp His Thr Gly Asp Ile Pro Cys Ile Cys Lys Glu 325 330 335			1362
TGT GGG AAG GTC TTT ATT CAG AGA TCA GAA TTG CTT ACG CAC CAG AAA Cys Gly Lys Val Phe Ile Gln Arg Ser Glu Leu Leu Thr His Gln Lys 340 345 350 355			1410
ACA CAC ACT AGA AAG AAG CCC TAT AAA TGC CAT GAC TGT GGA AAA GCC Thr His Thr Arg Lys Lys Pro Tyr Lys Cys His Asp Cys Gly Lys Ala 360 365 370			1458
TTT TTC CAG ATG TTA TCT CTC TTC AGA CAT CAG AGA ACT CAC AGT AGA Phe Phe Gln Met Leu Ser Leu Phe Arg His Gln Arg Thr His Ser Arg 375 380 385			1506
GAA AAA CTC TAT GAA TGC AGT GAA TGT GGC AAA GGC TTC TCC CAA AAC Glu Lys Leu Tyr Glu Cys Ser Glu Cys Gly Lys Gly Phe Ser Gln Asn 390 395 400			1554
TCA ACC CTC ATT ATA CAT CAG AAA ATT CAT ACT GGT GAG AGA CAG TAT Ser Thr Leu Ile Ile His Gln Lys Ile His Thr Gly Glu Arg Gln Tyr 405 410 415			1602
GCA TGC AGT GAA TGT GGG AAA GCC TTT ACC CAG AAG TCA ACA CTC AGC Ala Cys Ser Glu Cys Gly Lys Ala Phe Thr Gln Lys Ser Thr Leu Ser 420 425 430 435			1650
TTG CAC CAG AGA ATC CAC TCA GGG CAG AAG TCC TAT GIG TGT ATC GAA Leu His Gln Arg Ile His Ser Gly Gln Lys Ser Tyr Val Cys Ile Glu 440 445 450			1698
TGC GGG CAG GCC TTC ATC CAG AAG GCA CAC CTG ATT GTC CAT CAA AGA Cys Gly Gln Ala Phe Ile Gln Lys Ala His Leu Ile Val His Gln Arg 455 460 465			1746
AGC CAC ACA GGA GAA AAA CCT TAT CAG TGC CAC AAC TGT GGG AAA TCC Ser His Thr Gly Glu Lys Pro Tyr Gln Cys His Asn Cys Gly Lys Ser 470 475 480			1794

TTC ATT TCC AAG TCA CAG CTT GAT ATA CAT CAT CGA ATT CAT ACA GGG Phe Ile Ser Lys Ser Gln Leu Asp Ile His His Arg Ile His Thr Gly 485 490 495	1842
GAG AAA CCT TAT GAA TGC AGT GAC TGT GGA AAA ACC TTC ACC CAA AAG Glu Lys Pro Tyr Glu Cys Ser Asp Cys Gly Lys Thr Phe Thr Gln Lys 500 505 510 515	1890
TCA CAC CTG AAT ATA CAC CAG AAA ATT CAT ACT GGA GAA AGA CAC CAT Ser His Leu Asn Ile His Gln Lys Ile His Thr Gly Glu Arg His His 520 525 530	1938
GTA TGC AGT GAA TGC GGG AAA GCC TTC AAC CAG AAG TCA ATA CTC AGC Val Cys Ser Glu Cys Gly Lys Ala Phe Asn Gln Lys Ser Ile Leu Ser 535 540 545	1986
ATG CAT CAG AGA ATT CAC ACC GGA GAG AAG CCT TAC AAA TGC AGT GAA Met His Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Ser Glu 550 555 560	2034
TGT GGG AAA GCC TTC ACT TCT AAG TCT CAA TTC AAA GAG CAT CAG CGA Cys Gly Lys Ala Phe Thr Ser Lys Ser Gln Phe Lys Glu His Gln Arg 565 570 575	2082
ATT CAC ACG GGT GAG AAA CCC TAT GTG TGC ACT GAA TGT GGG AAG GCC Ile His Thr Gly Glu Lys Pro Tyr Val Cys Thr Glu Cys Gly Lys Ala 580 585 590 595	2130
TTC AAC GGC AGG TCA AAT TTC CAT AAA CAT CAA ATA ACT CAC ACT AGA Phe Asn Gly Arg Ser Asn Phe His Lys His Gln Ile Thr His Thr Arg 600 605 610	2178
GAG AGG CCT TTT GTC TGT TAC AAA TGT GGG AAG GCT TTT GTC CAG AAA Glu Arg Pro Phe Val Cys Tyr Lys Cys Gly Lys Ala Phe Val Gln Lys 615 620 625	2226
TCA GAG TTG ATT ACC CAT CAA AGA ACT CAC ATG GGA GAG AAA CCC TAT Ser Glu Leu Ile Thr His Gln Arg Thr His Met Gly Glu Lys Pro Tyr 630 635 640	2274
GAA TGC CTT GAC TGT GGG AAA TCG TTC AGT AAG AAA CCA CAA CTC AAG Glu Cys Leu Asp Cys Gly Lys Ser Phe Ser Lys Lys Pro Gln Leu Lys 645 650 655	2322
GTG CAT CAG CGA ATT CAC ACG GGA GAA AGA CCT TAT GTG TGT TCT GAA Val His Gln Arg Ile His Thr Gly Glu Arg Pro Tyr Val Cys Ser Glu 660 665 670 675	2370
TGT GGA AAG GCC TTC AAC AAC AGG TCA AAC TTC AAT AAA CAC CAA ACA	2418

Cys	Gly	Lys	Ala	Phe	Asn	Asn	Arg	Ser	Asn	Phe	Asn	Lys	His	Gln	Thr			
				680				685				690						
ACT	CAT	ACC	AGA	GAC	AAA	TCT	TAC	AAA	TGC	AGT	TAT	TCT	GTG	AAA	GGC	2466		
Thr	His	Thr	Arg	Asp	Lys	Ser	Tyr	Lys	Cys	Ser	Tyr	Ser	Val	Lys	Gly			
				695				700				705						
TTT	ACC	AAG	CAA	TGA	ATTCCTA	GTGCATCAGC	ATATTCATAA	ATGAAATATA								2518		
Phe	Thr	Lys	Gln															
				710														
CTCCGAGTTT	CTTGAAGAAG	AGAACATCTT	CTCAGAATCA	GGTCTAATTA	TATGTTATTG											2578		
AATTCATGCT	TCAGAAAAAC	TCTAGGGATG	CACIGCATGT	GTGAACACAT	GATAAAAAAG											2638		
TCATGCTTTA	TTTTAGTGAG	GGCAATTACA	GAGAAAAGAG	TAAGCAGAAA	TGTCCTTCTG											2698		
AGTACTGGCC	TCATTAAGGA	TTATAAATTT	TCTCCCCGGG	AAGAAACCCT	GACTAACGCA											2758		
TTGAGAAAAG	CCTTTCGTGA	AAGAATGGTA	CAAGACAGGT	TGTTACTCGA	TTATTTATAG											2818		
TAAATATGT	GGGAAATTAT	ATCAATGATA	ACCCTGTTTA	TTGTGGGATA	TCAATATTTT											2878		
TAAAGTGCCA	ACACAGTCAT	GATAGGACAA	TATTTTATGT	GTGTGTGTGC	GCCTTATGTA											2938		
TATAAGCATA	TATATAATAT	ATAAGCATAT	TATTATATAC	AGGTGAGTA	TCCCTTCTCC											2998		
AAAATGCCTG	GGATCAGAAG	CATTTTGGAT	TTCAGATACT	TACAGATTTT	GGAATATTTG											3058		
CATTATATTT	ATTGGTTGAG	CATCCCTAAT	CTGAAAATCC	AAGATTAAAT	GCTCCAATTA											3118		
GCATTTCCCTT	TGAGCGTCAT	GTTAGAGTTC	AAAAAGTTTC	AGATTTTGGG	TTTTTCAGATT											3178		
AGGAATACCC	AACCTGTATG	TACGTATATT	TCTGTATCTA	TGTATGTATA	TATATGCATA											3238		
TGCAGACATA	TGTATATGGT	CTGGTCAGCA	TATGTGTATG	TATGCGTATG	TATGTATGTA											3298		
TGTATGCCCT	CAGTGCAGTG	GGGTTTGCTG	CAGAATTCAC	TGCATAGCAG	GAGATGTAAG											3358		
CAGATGAGTT	ATTTTTTAAG	AGAATCTAAT	CTAATTGTTT	TTATAAAAAT	TATTCCTTAT											3418		
TGAATATTTA	TATAATGAGG	TTGTATCAAC	AATGATTAAAC	TCCTTTATTA	TACATACACA											3478		
TGAATGTGCA	TTTTTTGGTAA	ATGCATAAAT	GAGATTCTAT	AATGTTTACT	GATCTTTTATA											3538		
TTACAGATTT	TCTCTTCTTT	TAGGATTAGC	TCAGCTTGCC	CCCCCTTTCC	ATCTCCACCA											3598		
TCTATAGTGA	GCCTCTCCAT	AATTAGTGCC	AACCATTAGT	CTCGTTTATA	TTTTTACACC											3658		

AGGAGTCAAC AAAGTGTGCC ATGGCCAAA TATGGCCTCC CAACTGTTTT TTAAAAATAA 3718
AGTTTTATTG GAACACAAAA AAAAAAAAAA AAAAAA 3754

[0324]

SEQ ID NO:13

SEQUENCE CHARACTERISTICS:

LENGTH: 389 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met Ala Asp Pro Arg Asp Lys Ala Leu Gln Asp Tyr Arg Lys Lys Leu
1 5 10 15
Leu Glu His Lys Glu Ile Asp Gly Arg Leu Lys Glu Leu Arg Glu Gln
20 25 30
Leu Lys Glu Leu Thr Lys Gln Tyr Glu Lys Ser Glu Asn Asp Leu Lys
35 40 45
Ala Leu Gln Ser Val Gly Gln Ile Val Gly Glu Val Leu Lys Gln Leu
50 55 60
Thr Glu Glu Lys Phe Ile Val Lys Ala Thr Asn Gly Pro Arg Tyr Val
65 70 75 80
Val Gly Cys Arg Arg Gln Leu Asp Lys Ser Lys Leu Lys Pro Gly Thr
85 90 95
Arg Val Ala Leu Asp Met Thr Thr Leu Thr Ile Met Arg Tyr Leu Pro
100 105 110
Arg Glu Val Asp Pro Leu Val Tyr Asn Met Ser His Glu Asp Pro Gly
115 120 125
Asn Val Ser Tyr Ser Glu Ile Gly Gly Leu Ser Glu Gln Ile Arg Glu
130 135 140
Leu Arg Glu Val Ile Glu Leu Pro Leu Thr Asn Pro Glu Leu Phe Gln

145		150		155		160
Arg Val Gly Ile Ile Pro Pro Lys Gly Cys Leu Leu Tyr Gly Pro Pro	165		170		175	
Gly Thr Gly Lys Thr Leu Leu Ala Arg Ala Val Ala Ser Gln Leu Asp	180		185		190	
Cys Asn Phe Leu Lys Val Val Ser Ser Ser Ile Val Asp Lys Tyr Ile	195		200		205	
Gly Glu Ser Ala Arg Leu Ile Arg Glu Met Phe Asn Tyr Ala Arg Asp	210		215		220	
His Gln Pro Cys Ile Ile Phe Met Asp Glu Ile Asp Ala Ile Gly Gly	225		230		235	240
Arg Arg Phe Ser Glu Gly Thr Ser Ala Asp Arg Glu Ile Gln Arg Thr	245		250		255	
Leu Met Glu Leu Leu Asn Gln Met Asp Gly Phe Asp Thr Leu His Arg	260		265		270	
Val Lys Met Thr Met Ala Thr Asn Arg Pro Asp Thr Leu Asp Pro Ala	275		280		285	
Leu Leu Arg Pro Gly Arg Leu Asp Arg Lys Ile His Ile Asp Leu Pro	290		295		300	
Asn Glu Gln Ala Arg Leu Asp Ile Leu Lys Ile His Ala Gly Pro Ile	305		310		315	320
Thr Lys His Gly Glu Ile Asp Tyr Glu Ala Ile Val Lys Leu Ser Asp	325		330		335	
Gly Phe Asn Gly Ala Asp Leu Arg Asn Val Cys Thr Glu Ala Gly Met	340		345		350	
Phe Ala Ile Arg Ala Asp His Asp Phe Val Val Gln Glu Asp Phe Met	355		360		365	
Lys Ala Val Arg Lys Val Ala Asp Ser Lys Lys Leu Glu Ser Lys Leu	370		375		380	
Asp Tyr Lys Pro Val	385					

SEQ ID NO:14

SEQUENCE CHARACTERISTICS:

LENGTH: 1167 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGGCGGACC CTAGAGATAA GCGCTTCAG GACTACCGCA AGAAGTTGCT TGAACACAAG	60
GAGATCGACG GCCGTCTTAA GGAGTTAAGG GAACAATTAA AAGAACTTAC CAAGCAGTAT	120
GAAAAGTCTG AAAATGATCT GAAGGCCCTA CAGAGTGTG GGCAGATCGT GGGTGAAGTG	180
CTTAAACAGT TAACTGAAGA AAAATTCATT GTTAAAGCTA CCAATGGACC AAGATATGTT	240
GTGGGTTGTC GTCGACAGCT TGACAAAAGT AAGCTGAAGC CAGGAACAAG AGTTGCITTG	300
GATATGACTA CACTAACTAT CATGAGATAT TTGCCGAGAG AGGTGGATCC ACTGGTTTAT	360
AACATGTCTC ATGAGGACCC TGGGAATGTT TCTTATTCTG AGATTGGAGG GCTATCAGAA	420
CAGATCCGGG AATTAAAGAGA GGTGATAGAA TTACCTCTTA CAAACCCAGA GTTATTTCAG	480
CGTGTAGGAA TAATACCTCC AAAAGGCTGT TTGTTATATG GACCACCAGG TACGGGAAAA	540
ACACTCTTGG CACGAGCCGT TGCTAGCCAG CTGGACTGCA ATTCTTTAAA GGTTGATCT	600
AGTTCTATTG TAGACAAGTA CATTGGTGAA AGTGCTCGTT TGATCAGAGA AATGTTTAAT	660
TATGCTAGAG ATCATCAACC ATGCATCATT TTTATGGATG AAATAGATGC TATTGGTGGT	720
CGTCGGTTTT CTGAGGGTAC TTCAGCTGAC AGAGAGATTC AGAGAACGTT AATGGAGTTA	780
CTGAATCAAA TGGATGGATT TGATACTCTG CATAGAGTTA AAATGACCAT GGCTACAAAC	840
AGACCAGATA CACTGGATCC TGCTTTGCTG CGTCCAGGAA GATTAGATAG AAAAATACAT	900
ATTGATTTGC CAAATGAACA AGCAAGATTA GACATACTGA AAATCCATGC AGGTCCCAT	960
ACAAAGCATG GTGAAATAGA TTATGAAGCA ATTGTGAAGC TTTCGGATGG CTTTAATGGA	1020

GCAGATCTGA GAAATGTTTG TACTGAAGCA GGTATGTTTCG CAATTCGTGC TGATCATGAT 1080
TTTGTAAGTAC AGGAAGACTT CATGAAAGCA GTCAGAAAAG TGGCTGATTC TAAGAAGCTG 1140
GAGTCTAAAT TGGACTACAA ACCTGTG 1167

[0326]

SEQ ID NO:15

SEQUENCE CHARACTERISTICS:

LENGTH: 1566 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-331G07

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 17..1183

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

GAGACGGCTT CTCATC ATG GCG GAC CCT AGA GAT AAG GCG CTT CAG GAC 49
Met Ala Asp Pro Arg Asp Lys Ala Leu Gln Asp
1 5 10
TAC CGC AAG AAG TTG CTT GAA CAC AAG GAG ATC GAC GGC CGT CTT AAG 97
Tyr Arg Lys Lys Leu Leu Glu His Lys Glu Ile Asp Gly Arg Leu Lys
15 20 25
GAG TTA AGG GAA CAA TTA AAA GAA CTT ACC AAG CAG TAT GAA AAG TCT 145
Glu Leu Arg Glu Gln Leu Lys Glu Leu Thr Lys Gln Tyr Glu Lys Ser

30	35	40	
GAA AAT GAT CTG AAG GCC CTA CAG AGT GTT GGG CAG ATC GTG GGT GAA Glu Asn Asp Leu Lys Ala Leu Gln Ser Val Gly Gln Ile Val Gly Glu 45 50 55			193
GTG CTT AAA CAG TTA ACT GAA GAA AAA TTC ATT GTT AAA GCT ACC AAT Val Leu Lys Gln Leu Thr Glu Glu Lys Phe Ile Val Lys Ala Thr Asn 60 65 70 75			241
GGA CCA AGA TAT GTT GTG GGT TGT CGT CGA CAG CTT GAC AAA AGT AAG Gly Pro Arg Tyr Val Val Gly Cys Arg Arg Gln Leu Asp Lys Ser Lys 80 85 90			289
CTG AAG CCA GGA ACA AGA GTT GCT TTG GAT ATG ACT ACA CTA ACT ATC Leu Lys Pro Gly Thr Arg Val Ala Leu Asp Met Thr Thr Leu Thr Ile 95 100 105			337
ATG AGA TAT TTG CCG AGA GAG GTG GAT CCA CTG GTT TAT AAC ATG TCT Met Arg Tyr Leu Pro Arg Glu Val Asp Pro Leu Val Tyr Asn Met Ser 110 115 120			385
CAT GAG GAC CCT GGG AAT GTT TCT TAT TCT GAG ATT GGA GGG CTA TCA His Glu Asp Pro Gly Asn Val Ser Tyr Ser Glu Ile Gly Gly Leu Ser 125 130 135			433
GAA CAG ATC CGG GAA TTA AGA GAG GTG ATA GAA TTA CCT CTT ACA AAC Glu Gln Ile Arg Glu Leu Arg Glu Val Ile Glu Leu Pro Leu Thr Asn 140 145 150 155			481
CCA GAG TTA TTT CAG CGT GTA GGA ATA ATA CCT CCA AAA GGC TGT TTG Pro Glu Leu Phe Gln Arg Val Gly Ile Ile Pro Pro Lys Gly Cys Leu 160 165 170			529
TTA TAT GGA CCA CCA GGT ACG GGA AAA ACA CTC TTG GCA CGA GCC GTT Leu Tyr Gly Pro Pro Gly Thr Gly Lys Thr Leu Leu Ala Arg Ala Val 175 180 185			577
GCT AGC CAG CTG GAC TGC AAT TTC TTA AAG GTT GTA TCT AGT TCT ATT Ala Ser Gln Leu Asp Cys Asn Phe Leu Lys Val Val Ser Ser Ser Ile 190 195 200			625
GTA GAC AAG TAC ATT GGT GAA AGT GCT CGT TTG ATC AGA GAA ATG TTT Val Asp Lys Tyr Ile Gly Glu Ser Ala Arg Leu Ile Arg Glu Met Phe 205 210 215			673
AAT TAT GCT AGA GAT CAT CAA CCA TGC ATC ATT TTT ATG GAT GAA ATA Asn Tyr Ala Arg Asp His Gln Pro Cys Ile Ile Phe Met Asp Glu Ile 220 225 230 235			721

GAT GCT ATT GGT GGT CGT CGG TTT TCT GAG GGT ACT TCA GCT GAC AGA Asp Ala Ile Gly Gly Arg Arg Phe Ser Glu Gly Thr Ser Ala Asp Arg 240 245 250	769
GAG ATT CAG AGA ACG TTA ATG GAG TTA CTG AAT CAA ATG GAT GGA TTT Glu Ile Gln Arg Thr Leu Met Glu Leu Leu Asn Gln Met Asp Gly Phe 255 260 265	817
GAT ACT CTG CAT AGA GTT AAA ATG ACC ATG GCT ACA AAC AGA CCA GAT Asp Thr Leu His Arg Val Lys Met Thr Met Ala Thr Asn Arg Pro Asp 270 275 280	865
ACA CTG GAT CCT GCT TTG CTG CGT CCA GGA AGA TTA GAT AGA AAA ATA Thr Leu Asp Pro Ala Leu Leu Arg Pro Gly Arg Leu Asp Arg Lys Ile 285 290 295	913
CAT ATT GAT TTG CCA AAT GAA CAA GCA AGA TTA GAC ATA CTG AAA ATC His Ile Asp Leu Pro Asn Glu Gln Ala Arg Leu Asp Ile Leu Lys Ile 300 305 310 315	961
CAT GCA GGT CCC ATT ACA AAG CAT GGT GAA ATA GAT TAT GAA GCA ATT His Ala Gly Pro Ile Thr Lys His Gly Glu Ile Asp Tyr Glu Ala Ile 320 325 330	1009
GIG AAG CTT TCG GAT GGC TTT AAT GGA GCA GAT CTG AGA AAT GTT TGT Val Lys Leu Ser Asp Gly Phe Asn Gly Ala Asp Leu Arg Asn Val Cys 335 340 345	1057
ACT GAA GCA GGT ATG TTC GCA ATT CGT GCT GAT CAT GAT TTT GTA GTA Thr Glu Ala Gly Met Phe Ala Ile Arg Ala Asp His Asp Phe Val Val 350 355 360	1105
CAG GAA GAC TTC ATG AAA GCA GTC AGA AAA GTG GCT GAT TCT AAG AAG Gln Glu Asp Phe Met Lys Ala Val Arg Lys Val Ala Asp Ser Lys Lys 365 370 375	1153
CTG GAG TCT AAA TTG GAC TAC AAA CCT GTG TAATTTACTG TAAGATTTTT Leu Glu Ser Lys Leu Asp Tyr Lys Pro Val 380 385	1203
GATGGCTGCA TGACAGATGT TGGCTTATTG TAAAAATAAA GTTAAAGAAA ATAATGTATG	1263
TATTGGCAAT GATGTCATTA AAAGTATATG AATAAAAATA TGAGTAACAT CATAAAAATT	1323
AGTAATTCAA CTTTTAAGAT ACAGAAGAAA TTTGTATGTT TGTAAAGTT GCATTTATTG	1383
CAGCAAGTTA CAAAGGGAAA GIGTTGAAGC TTTTCATATT TGCTGCGTGA GCATTTTGTA	1443
AAATATTGAA AGTGGTTTGA GATAGTGGTA TAAGAAAGCA TTTCTTATGA CTTATTTTGT	1503

ATCATTTGTT TTCCATCTCT AAAAAGTTGA ATAAAATCTG TTTGATTCAG TTCTCCTAAA 1563
AAA 1566

[0327]

SEQ ID NO:16

SEQUENCE CHARACTERISTICS:

LENGTH: 223 amino acids

SEQUENCE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met Ser Asp Glu Glu Ala Arg Gln Ser Gly Gly Ser Ser Gln Ala Gly
1 5 10 15
Val Val Thr Val Ser Asp Val Gln Glu Leu Met Arg Arg Lys Glu Glu
20 25 30
Ile Glu Ala Gln Ile Lys Ala Asn Tyr Asp Val Leu Glu Ser Gln Lys
35 40 45
Gly Ile Gly Met Asn Glu Pro Leu Val Asp Cys Glu Gly Tyr Pro Arg
50 55 60
Ser Asp Val Asp Leu Tyr Gln Val Arg Thr Ala Arg His Asn Ile Ile
65 70 75 80
Cys Leu Gln Asn Asp His Lys Ala Val Met Lys Gln Val Glu Glu Ala
85 90 95
Leu His Gln Leu His Ala Arg Asp Lys Glu Lys Gln Ala Arg Asp Met
100 105 110
Ala Glu Ala His Lys Glu Ala Met Ser Arg Lys Leu Gly Gln Ser Glu
115 120 125
Ser Gln Gly Pro Pro Arg Ala Phe Ala Lys Val Asn Ser Ile Ser Pro
130 135 140
Gly Ser Pro Ala Ser Ile Ala Gly Leu Gln Val Asp Asp Glu Ile Val

-133-

145					150					155					160
Glu	Phe	Gly	Ser	Val	Asn	Thr	Gln	Asn	Phe	Gln	Ser	Leu	His	Asn	Ile
				165					170					175	
Gly	Ser	Val	Val	Gln	His	Ser	Glu	Gly	Lys	Pro	Leu	Asn	Val	Thr	Val
			180					185					190		
Ile	Arg	Arg	Gly	Glu	Lys	His	Gln	Leu	Arg	Leu	Val	Pro	Thr	Arg	Trp
		195					200					205			
Ala	Gly	Lys	Gly	Leu	Leu	Gly	Cys	Asn	Ile	Ile	Pro	Leu	Gln	Arg	
	210					215					220				

[0328]

SEQ ID NO:17

SEQUENCE CHARACTERISTICS:

LENGTH: 669 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGTCCGACG AGGAAGCGAG GCAGAGCGGA GGCTCCTCGC AGGCCGGCGT CGTGACTGTC	60
AGCGACGTCC AGGAGCTGAT GCGGCGCAAG GAGGAGATAG AAGCGCAGAT CAAGGCCAAC	120
TATGACGTGC TGGAAAGCCA AAAAGGCATT GGGATGAACG AGCCGCTGGT GGACTGTGAG	180
GGCTACCCCC GGTCAGACGT GGACCTGTAC CAAGTCCGCA CCGCCAGGCA CAACATCATA	240
TGCCTGCAGA ATGATCACAA GGCAGTGATG AAGCAGGTGG AGGAGGCCCT GCACCAGCTG	300
CACGCTCGCG ACAAGGAGAA GCAGGCCCGG GACATGGCTG AGGCCACAA AGAGGCCATG	360
AGCCGCAAAC TGGGTCAGAG TGAGAGCCAG GGCCCTCCAC GGGCCTTCGC CAAAGTGAAC	420
AGCATCAGCC CCGGCTCCCC AGCCAGCATC GCGGGTCTGC AAGTGGATGA TGAGATTGTG	480

-134-

GAGTTCGGCT CTGTGAACAC CCAGAACTTC CAGTCACTGC ATAACATTGG CAGTGTGGTG 540
CAGCACAGTG AGGGGAAGCC CCTGAATGTG ACAGTGATCC GCAGGGGGGA AAAACACCAG 600
CTTAGACTTG TTCCAACACG CTGGGCAGGA AAAGGACTGC TGGGCTGCAA CATTATTCCT 660
CTGCAAAGA 669

[0329]

SEQ ID NO:18

SEQUENCE CHARACTERISTICS:

LENGTH: 1128 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-163D09

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 125..793

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

ACTGTTCTCG CGTTCGCGGA CGGCTGTGGT GTTTTGGCGC ATGGGCGGAG CGTAGTTACG 60
GTCGACTGGG GCGTCGTCCC TAGCCCGGGA GCCGGGTCTC TGGAGTCGCG GCCCGGGGTT 120
CACG ATG TCC GAC GAG GAA GCG AGG CAG AGC GGA GGC TCC TCG CAG GCC 169
Met Ser Asp Glu Glu Ala Arg Gln Ser Gly Gly Ser Ser Gln Ala
1 5 10 15

GGC GTC GTG ACT GTC AGC GAC GTC CAG GAG CTG ATG CGG CGC AAG GAG Gly Val Val Thr Val Ser Asp Val Gln Glu Leu Met Arg Arg Lys Glu 20 25 30	217
GAG ATA GAA GCG CAG ATC AAG GCC AAC TAT GAC GTG CTG GAA AGC CAA Glu Ile Glu Ala Gln Ile Lys Ala Asn Tyr Asp Val Leu Glu Ser Gln 35 40 45	265
AAA GGC ATT GGG ATG AAC GAG CCG CTG GTG GAC TGT GAG GGC TAC CCC Lys Gly Ile Gly Met Asn Glu Pro Leu Val Asp Cys Glu Gly Tyr Pro 50 55 60	313
CGG TCA GAC GTG GAC CTG TAC CAA GTC CGC ACC GCC AGG CAC AAC ATC Arg Ser Asp Val Asp Leu Tyr Gln Val Arg Thr Ala Arg His Asn Ile 65 70 75	361
ATA TGC CTG CAG AAT GAT CAC AAG GCA GTG ATG AAG CAG GTG GAG GAG Ile Cys Leu Gln Asn Asp His Lys Ala Val Met Lys Gln Val Glu Glu 80 85 90 95	409
GCC CTG CAC CAG CTG CAC GCT CGC GAC AAG GAG AAG CAG GCC CGG GAC Ala Leu His Gln Leu His Ala Arg Asp Lys Glu Lys Gln Ala Arg Asp 100 105 110	457
ATG GCT GAG GCC CAC AAA GAG GCC ATG AGC CGC AAA CTG GGT CAG AGT Met Ala Glu Ala His Lys Glu Ala Met Ser Arg Lys Leu Gly Gln Ser 115 120 125	505
GAG AGC CAG GGC CCT CCA CGG GCC TTC GCC AAA GTG AAC AGC ATC AGC Glu Ser Gln Gly Pro Pro Arg Ala Phe Ala Lys Val Asn Ser Ile Ser 130 135 140	553
CCC GGC TCC CCA GCC AGC ATC GCG GGT CTG CAA GTG GAT GAT GAG ATT Pro Gly Ser Pro Ala Ser Ile Ala Gly Leu Gln Val Asp Asp Glu Ile 145 150 155	601
GTG GAG TTC GGC TCT GTG AAC ACC CAG AAC TTC CAG TCA CTG CAT AAC Val Glu Phe Gly Ser Val Asn Thr Gln Asn Phe Gln Ser Leu His Asn 160 165 170 175	649
ATT GGC AGT GTG GTG CAG CAC AGT GAG GGG AAG CCC CTG AAT GTG ACA Ile Gly Ser Val Val Gln His Ser Glu Gly Lys Pro Leu Asn Val Thr 180 185 190	697
GTG ATC CGC AGG GGG GAA AAA CAC CAG CTT AGA CTT GTT CCA ACA CGC Val Ile Arg Arg Gly Glu Lys His Gln Leu Arg Leu Val Pro Thr Arg 195 200 205	745
TGG GCA GGA AAA GGA CTG CTG GGC TGC AAC ATT ATT CCT CTG CAA AGA	793

Trp Ala Gly Lys Gly Leu Leu Gly Cys Asn Ile Ile Pro Leu Gln Arg
210 215 220

TGATTGTCCC TGGGGAACAG TAACAGGAAA GCATCTTCCC TTGCCCTGGA CTTGGGTCTA 853
GGGATTTC CA ACTTGCTCTC TCTCCCTGAA GCATAAGGAT CTGGAAGAGG CTTGTAACT 913
GAACTTCTGT GTGGTGGCAG TACTGTGGCC CACCAGTGTA ATCTCCCTGG ATTAAGGCAT 973
TCTTAAAAAC TTAGGCTTGG CCTCTTTTAC AAATTAGGCC ACGGCCCTAA ATAGGAATTC 1033
CCTGGATTGT GGGCAAGTGG GCGGAAGTTA TTCTGGCAGG TACTGGTGTG ATTATTATTA 1093
TTATTTTAA TAAAGAGTTT TACAGTGCTG ATATG 1128

[0330]

SEQ ID NO:19

SEQUENCE CHARACTERISTICS:

LENGTH: 506 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met Ala Glu Ala Asp Phe Lys Met Val Ser Glu Pro Val Ala His Gly
1 5 10 15
Val Ala Glu Glu Glu Met Ala Ser Ser Thr Ser Asp Ser Gly Glu Glu
20 25 30
Ser Asp Ser Ser Ser Ser Ser Ser Ser Thr Ser Asp Ser Ser Ser Ser
35 40 45
Ser Ser Thr Ser Gly Ser Ser Ser Gly Ser Gly Ser Ser Ser Ser Ser
50 55 60
Ser Gly Ser Thr Ser Ser Arg Ser Arg Leu Tyr Arg Lys Lys Arg Val
65 70 75 80
Pro Glu Pro Ser Arg Arg Ala Arg Arg Ala Pro Leu Gly Thr Asn Phe
85 90 95

Val Asp Arg Leu Pro Gln Ala Val Arg Asn Arg Val Gln Ala Leu Arg
100 105 110

Asn Ile Gln Asp Glu Cys Asp Lys Val Asp Thr Leu Phe Leu Lys Ala
115 120 125

Ile His Asp Leu Glu Arg Lys Tyr Ala Glu Leu Asn Lys Pro Leu Tyr
130 135 140

Asp Arg Arg Phe Gln Ile Ile Asn Ala Glu Tyr Glu Pro Thr Glu Glu
145 150 155 160

Glu Cys Glu Trp Asn Ser Glu Asp Glu Glu Phe Ser Ser Asp Glu Glu
165 170 175

Val Gln Asp Asn Thr Pro Ser Glu Met Pro Pro Leu Glu Gly Glu Glu
180 185 190

Glu Glu Asn Pro Lys Glu Asn Pro Glu Val Lys Ala Glu Glu Lys Glu
195 200 205

Val Pro Lys Glu Ile Pro Glu Val Lys Asp Glu Glu Lys Glu Val Ala
210 215 220

Lys Glu Ile Pro Glu Val Lys Ala Glu Glu Lys Ala Asp Ser Lys Asp
225 230 235 240

Cys Met Glu Ala Thr Pro Glu Val Lys Glu Asp Pro Lys Glu Val Pro
245 250 255

Gln Val Lys Ala Asp Asp Lys Glu Gln Pro Lys Ala Thr Glu Ala Lys
260 265 270

Ala Arg Ala Ala Val Arg Glu Thr His Lys Arg Val Pro Glu Glu Arg
275 280 285

Leu Arg Asp Ser Val Asp Leu Lys Arg Ala Arg Lys Gly Lys Pro Lys
290 295 300

Arg Glu Asp Pro Lys Gly Ile Pro Asp Tyr Trp Leu Ile Val Leu Lys
305 310 315 320

Asn Val Asp Lys Leu Gly Pro Met Ile Gln Lys Tyr Asp Glu Pro Ile
325 330 335

Leu Lys Phe Leu Ser Asp Val Ser Leu Lys Phe Ser Lys Pro Gly Gln
340 345 350

Pro Val Ser Tyr Thr Phe Glu Phe His Phe Leu Pro Asn Pro Tyr Phe

-138-

355		360		365											
Arg	Asn	Glu	Val	Leu	Val	Lys	Thr	Tyr	Ile	Ile	Lys	Ala	Lys	Pro	Asp
370						375					380				
His	Asn	Asp	Pro	Phe	Phe	Ser	Trp	Gly	Trp	Glu	Ile	Glu	Asp	Cys	Lys
385					390					395					400
Gly	Cys	Lys	Ile	Asp	Arg	Arg	Arg	Gly	Lys	Asp	Val	Thr	Val	Thr	Thr
				405					410					415	
Thr	Gln	Ser	Arg	Thr	Thr	Ala	Thr	Gly	Glu	Ile	Glu	Ile	Gln	Pro	Arg
			420					425					430		
Val	Val	Pro	Asn	Ala	Ser	Phe	Phe	Asn	Phe	Phe	Ser	Pro	Pro	Glu	Ile
		435					440					445			
Pro	Met	Ile	Gly	Lys	Leu	Glu	Pro	Arg	Glu	Asp	Ala	Ile	Leu	Asp	Glu
	450					455					460				
Asp	Phe	Glu	Ile	Gly	Gln	Ile	Leu	His	Asp	Asn	Val	Ile	Leu	Lys	Ser
465					470					475					480
Ile	Tyr	Tyr	Tyr	Thr	Gly	Glu	Val	Asn	Gly	Thr	Tyr	Tyr	Gln	Phe	Gly
				485					490					495	
Lys	His	Tyr	Gly	Asn	Lys	Lys	Tyr	Arg	Lys						
			500					505							

[0331]

SEQ ID NO:20

SEQUENCE CHARACTERISTICS:

LENGTH: 1518 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGGCAGAAG CAGATTTTAA AATGGTCTCG GAACCTGTCTG CCCATGGGGT TGCCGAAGAG

GAGATGGCTA GCTCGACTAG TGATTCTGGG GAAGAATCTG ACAGCAGTAG CTCTAGCAGC	120
AGCACTAGTG ACAGCAGCAG CAGCAGCAGC ACTAGTGGCA GCAGCAGCGG CAGCGGCAGC	180
AGCAGCAGCA GCAGCGGCAG CACTAGCAGC CGCAGCCGCT TGTATAGAAA GAAGAGGGTA	240
CCTGAGCCTT CCAGAAGGGC GCGGCGGGCC CCGTTGGGAA CAAATTTTCGT GGATAGGCTG	300
CCTCAGGCAG TTAGAAATCG TGTGCAAGCG CTTAGAAACA TTCAAGATGA ATGTGACAAG	360
GTAGATACCC TGTTCITAAA AGCAATTCAT GATCTTGAAA GAAAATATGC TGAACTCAAC	420
AAGCCTCTGT ATGATAGGCG GTTTCAAATC ATCAATGCAG AATACGAGCC TACAGAAGAA	480
GAATGTGAAT GGAATTCAGA GGATGAGGAG TTCAGCAGTG ATGAGGAGGT GCAGGATAAC	540
ACCCCTAGTG AAATGCCTCC CTTAGAGGGT GAGGAAGAAG AAAACCCTAA AGAAAACCCA	600
GAGGTGAAAG CTGAAGAGAA GGAAGTTCCT AAAGAAATTC CTGAGGTGAA GGATGAAGAA	660
AAGGAAGTTG CTAAAGAAAT TCCTGAGGTA AAGGCTGAAG AAAAAGCAGA TTCTAAAGAC	720
TGTATGGAGG CAACCCCTGA AGTAAAAGAA GATCCTAAAG AAGTCCCCCA GGTAAAGGCA	780
GATGATAAAG AACAGCCTAA AGCAACAGAG GCTAAGGCAA GGGCTGCAGT AAGAGAGACT	840
CATAAAAGAG TTCCTGAGGA AAGGCTTCGG GACAGTGTAG ATCTTAAAG AGCTAGGAAG	900
GGAAAGCCTA AAAGAGAAGA CCCTAAAGGC ATTCTGACT ATTGGCTGAT TGTTTTAAAG	960
AATGTTGACA AGCTCGGGCC TATGATTCAG AAGTATGATG AGCCCATTCCT GAAGTCTTG	1020
TCGGATGTTA GCCTGAAGTT CTCAAAACCT GGCCAGCCTG TAAGTTACAC CTTTGAATTT	1080
CATTTTCTAC CCAACCCATA CTTCAGAAAT GAGGTGCTGG TGAAGACATA TATAATAAAG	1140
GCAAAACCAG ATCACAATGA TCCCTTCITT TCTTGGGGAT GGGAAATGA AGATTGCAAA	1200
GGCTGCAAGA TAGACCGGAG AAGAGGAAAA GATGTTACTG TGACAACTAC CCAGAGTCGC	1260
ACAACCTGCTA CTGGAGAAAT TGAAATCCAG CCAAGAGTGG TTCCTAATGC ATCATTCTTC	1320
AACTTCTTTA GTCTCTCTGA GATTCCTATG ATTGGGAAGC TGAACCACG AGAAGATGCT	1380
ATCCTGGATG AGGACTTTGA AATTGGGCAG ATTTTACATG ATAATGTCAT CCTGAAATCA	1440
ATCTATTACT ATACTGGAGA AGTCAATGGT ACCTACTATC AATTTGGCAA ACATTATGGA	1500
AACAAGAAAT ACAGAAAA	1518

[0332]

SEQ ID NO:21

SEQUENCE CHARACTERISTICS:

LENGTH: 2636 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-078D05

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 266..1783

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

GATTCGGCTG CGGTACATCT CGGCACTCTA GCTGCAGCCG GGAGAGGCCT TGCCGCCACC	60
GCTGTGCGCC AAGCCTCCAC TGCCGCTGCC ACCTCAGCGC CGGCCTCTGC ATCCCCAGCT	120
CCAGCTCCGC TCTGCGCCGC TGCTGCCATC GCCGCTGCCA CCTCCGCAGC CCGGGCCTCC	180
GCCGCCGCCA CCCAAGCATC CGTGAGTCAT TTTCTGCCCA TCTCTGGTCG CGCGGTCTCC	240
CTGGTAGAGT TTGTAGGCTT GCAAG ATG GCA GAA GCA GAT TTT AAA ATG GTC	292
Met Ala Glu Ala Asp Phe Lys Met Val	
1 5	
TCG GAA CCT GTC GCC CAT GGG GTT GCC GAA GAG GAG ATG GCT AGC TCG	340
Ser Glu Pro Val Ala His Gly Val Ala Glu Glu Glu Met Ala Ser Ser	
10 15 20 25	

ACT	AGT	GAT	TCT	GGG	GAA	GAA	TCT	GAC	AGC	AGT	AGC	TCT	AGC	AGC	AGC	388
Thr	Ser	Asp	Ser	Gly	Glu	Glu	Ser	Asp	Ser	Ser	Ser	Ser	Ser	Ser	Ser	
				30					35					40		
ACT	AGT	GAC	AGC	AGC	AGC	AGC	AGC	AGC	ACT	AGT	GGC	AGC	AGC	AGC	GGC	436
Thr	Ser	Asp	Ser	Ser	Ser	Ser	Ser	Ser	Thr	Ser	Gly	Ser	Ser	Ser	Gly	
			45					50					55			
AGC	GGC	AGC	AGC	AGC	AGC	AGC	AGC	GGC	AGC	ACT	AGC	AGC	CGC	AGC	CGC	484
Ser	Gly	Ser	Ser	Ser	Ser	Ser	Ser	Gly	Ser	Thr	Ser	Ser	Arg	Ser	Arg	
		60					65					70				
TTG	TAT	AGA	AAG	AAG	AGG	GTA	CCT	GAG	CCT	TCC	AGA	AGG	GCG	CGG	CGG	532
Leu	Tyr	Arg	Lys	Lys	Arg	Val	Pro	Glu	Pro	Ser	Arg	Arg	Ala	Arg	Arg	
	75					80					85					
GCC	CCG	TTG	GGA	ACA	AAT	TTC	GTG	GAT	AGG	CTG	CCT	CAG	GCA	GTT	AGA	580
Ala	Pro	Leu	Gly	Thr	Asn	Phe	Val	Asp	Arg	Leu	Pro	Gln	Ala	Val	Arg	
90					95					100					105	
AAT	CGT	GTG	CAA	GCG	CTT	AGA	AAC	ATT	CAA	GAT	GAA	TGT	GAC	AAG	GTA	628
Asn	Arg	Val	Gln	Ala	Leu	Arg	Asn	Ile	Gln	Asp	Glu	Cys	Asp	Lys	Val	
			110					115						120		
GAT	ACC	CTG	TTC	TTA	AAA	GCA	ATT	CAT	GAT	CTT	GAA	AGA	AAA	TAT	GCT	676
Asp	Thr	Leu	Phe	Leu	Lys	Ala	Ile	His	Asp	Leu	Glu	Arg	Lys	Tyr	Ala	
			125					130					135			
GAA	CTC	AAC	AAG	CCT	CTG	TAT	GAT	AGG	CGG	TTT	CAA	ATC	ATC	AAT	GCA	724
Glu	Leu	Asn	Lys	Pro	Leu	Tyr	Asp	Arg	Arg	Phe	Gln	Ile	Ile	Asn	Ala	
		140					145					150				
GAA	TAC	GAG	CCT	ACA	GAA	GAA	GAA	TGT	GAA	TGG	AAT	TCA	GAG	GAT	GAG	772
Glu	Tyr	Glu	Pro	Thr	Glu	Glu	Glu	Cys	Glu	Trp	Asn	Ser	Glu	Asp	Glu	
	155				160						165					
GAG	TTC	AGC	AGT	GAT	GAG	GAG	GTG	CAG	GAT	AAC	ACC	CCT	AGT	GAA	ATG	820
Glu	Phe	Ser	Ser	Asp	Glu	Glu	Val	Gln	Asp	Asn	Thr	Pro	Ser	Glu	Met	
170					175					180					185	
CCT	CCC	TTA	GAG	GGT	GAG	GAA	GAA	GAA	AAC	CCT	AAA	GAA	AAC	CCA	GAG	868
Pro	Pro	Leu	Glu	Gly	Glu	Glu	Glu	Glu	Asn	Pro	Lys	Glu	Asn	Pro	Glu	
				190					195					200		
GTG	AAA	GCT	GAA	GAG	AAG	GAA	GTT	CCT	AAA	GAA	ATT	CCT	GAG	GTG	AAG	916
Val	Lys	Ala	Glu	Glu	Lys	Glu	Val	Pro	Lys	Glu	Ile	Pro	Glu	Val	Lys	
			205					210					215			
GAT	GAA	GAA	AAG	GAA	GTT	GCT	AAA	GAA	ATT	CCT	GAG	GTA	AAG	GCT	GAA	964

Asp Glu Glu Lys Glu Val Ala Lys Glu Ile Pro Glu Val Lys Ala Glu	
220 225 230	
GAA AAA GCA GAT TCT AAA GAC TGT ATG GAG GCA ACC CCT GAA GTA AAA	1012
Glu Lys Ala Asp Ser Lys Asp Cys Met Glu Ala Thr Pro Glu Val Lys	
235 240 245	
GAA GAT CCT AAA GAA GTC CCC CAG GTA AAG GCA GAT GAT AAA GAA CAG	1060
Glu Asp Pro Lys Glu Val Pro Gln Val Lys Ala Asp Asp Lys Glu Gln	
250 255 260 265	
CCT AAA GCA ACA GAG GCT AAG GCA AGG GCT GCA GTA AGA GAG ACT CAT	1108
Pro Lys Ala Thr Glu Ala Lys Ala Arg Ala Ala Val Arg Glu Thr His	
270 275 280	
AAA AGA GTT CCT GAG GAA AGG CTT CGG GAC AGT GTA GAT CTT AAA AGA	1156
Lys Arg Val Pro Glu Glu Arg Leu Arg Asp Ser Val Asp Leu Lys Arg	
285 290 295	
GCT AGG AAG GGA AAG CCT AAA AGA GAA GAC CCT AAA GGC ATT CCT GAC	1204
Ala Arg Lys Gly Lys Pro Lys Arg Glu Asp Pro Lys Gly Ile Pro Asp	
300 305 310	
TAT TGG CTG ATT GTT TTA AAG AAT GTT GAC AAG CTC GGG CCT ATG ATT	1252
Tyr Trp Leu Ile Val Leu Lys Asn Val Asp Lys Leu Gly Pro Met Ile	
315 320 325	
CAG AAG TAT GAT GAG CCC ATT CTG AAG TTC TTG TCG GAT GTT AGC CTG	1300
Gln Lys Tyr Asp Glu Pro Ile Leu Lys Phe Leu Ser Asp Val Ser Leu	
330 335 340 345	
AAG TTC TCA AAA CCT GGC CAG CCT GTA AGT TAC ACC TTT GAA TTT CAT	1348
Lys Phe Ser Lys Pro Gly Gln Pro Val Ser Tyr Thr Phe Glu Phe His	
350 355 360	
TTT CTA CCC AAC CCA TAC TTC AGA AAT GAG GTG CTG GTG AAG ACA TAT	1396
Phe Leu Pro Asn Pro Tyr Phe Arg Asn Glu Val Leu Val Lys Thr Tyr	
365 370 375	
ATA ATA AAG GCA AAA CCA GAT CAC AAT GAT CCC TTC TTT TCT TGG GGA	1444
Ile Ile Lys Ala Lys Pro Asp His Asn Asp Pro Phe Phe Ser Trp Gly	
380 385 390	
TGG GAA ATT GAA GAT TGC AAA GGC TGC AAG ATA GAC CGG AGA AGA GGA	1492
Trp Glu Ile Glu Asp Cys Lys Gly Cys Lys Ile Asp Arg Arg Arg Gly	
395 400 405	
AAA GAT GTT ACT GTG ACA ACT ACC CAG AGT CGC ACA ACT GCT ACT GGA	1540
Lys Asp Val Thr Val Thr Thr Thr Gln Ser Arg Thr Thr Ala Thr Gly	

410	415	420	425	
GAA ATT GAA ATC CAG CCA AGA GTG GTT CCT AAT GCA TCA TTC TTC AAC				1588
Glu Ile Glu Ile Gln Pro Arg Val Val Pro Asn Ala Ser Phe Phe Asn	430	435	440	
TTC TTT AGT CCT CCT GAG ATT CCT ATG ATT GGG AAG CTG GAA CCA CGA				1636
Phe Phe Ser Pro Pro Glu Ile Pro Met Ile Gly Lys Leu Glu Pro Arg	445	450	455	
GAA GAT GCT ATC CTG GAT GAG GAC TTT GAA ATT GGG CAG ATT TTA CAT				1684
Glu Asp Ala Ile Leu Asp Glu Asp Phe Glu Ile Gly Gln Ile Leu His	460	465	470	
GAT AAT GTC ATC CTG AAA TCA ATC TAT TAC TAT ACT GGA GAA GTC AAT				1732
Asp Asn Val Ile Leu Lys Ser Ile Tyr Tyr Tyr Thr Gly Glu Val Asn	475	480	485	
GGT ACC TAC TAT CAA TTT GGC AAA CAT TAT GGA AAC AAG AAA TAC AGA				1780
Gly Thr Tyr Tyr Gln Phe Gly Lys His Tyr Gly Asn Lys Lys Tyr Arg	490	495	500	505
AAA TAAGTCAATC TGAAAGATTT TTCAAGAATC TTAAAATCTC AAGAAGTGAA				1833
Lys				
GCAGATTCAT ACAGCCTTGA AAAAAGTAAA ACCCTGACCT GTAACCTGAA CACTATTATT				1893
CCTTATAGTC AAGTTTTTGT GGTTCCTTGG TAGTCTATAT TTTAAAAATA GTCCTAAAAA				1953
GTGCTTAAGT GCCAGTTTAT TCTATCTAGG CTGTTGTAGT ATAATATTCT TCAAAATATG				2013
TAAGCTGTTG TCAATTATCT AAAGCATGTT AGTTTGGTGC TACACAGTGT TGATTTTTGT				2073
GATGTCCTTT GGICATGTTT CTGTTAGACT GTAGCTGTGA AACTGTCAGA ATTGTTAACT				2133
GAAACAAATA TTGCTTGAA AAAAAAGIT CATGAAGTAC CAATGCAAGT GTTTTATTTT				2193
TTTTCTTTTT TCCAGCCCAT AAGACTAAGG GTTTAAATCT GCTTGCACTA GCTGTGCCTT				2253
CATTAGTTTG CTATAGAAAT CCAGTACTTA TAGTAAATAA AACAGTGTAT TTTGAAGTTT				2313
GACTGCTTGA AAAAGATTAG CACATCTA ATGTGAAAAG ACCACATTTG ATTCAACTGA				2373
GACCTTGTGT ATGTGACATA TAGTGGCCTA TAAATTTAAT CATAATGATG TTATTGTTTA				2433
CCACTGAGGT GTTAATATAA CATAGTATTT TTGAAAAAGT TTCTTCATCT TATATTGTGT				2493
AATTGTAAAC TAAAGATACC GTGTTTTCTT TGTATTGTTT TCTACCTTCC CTTTCACTGA				2553

AAATGATCAC TTCATTGAT ACTGTTTTTC ATGTTCTTGT ATTGCAACCT AAAATAAATA 2613
AATATTAAAG TGTGTTATAC TAT 2636

[0333]

SEQ ID NO:22

SEQUENCE CHARACTERISTICS:

LENGTH: 170 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met Thr Glu Leu Gln Ser Ala Leu Leu Leu Arg Arg Gln Leu Ala Glu
1 5 10 15
Leu Asn Lys Asn Pro Val Glu Gly Phe Ser Ala Gly Leu Ile Asp Asp
20 25 30
Asn Asp Leu Tyr Arg Trp Glu Val Leu Ile Ile Gly Pro Pro Asp Thr
35 40 45
Leu Tyr Glu Gly Gly Val Phe Lys Ala His Leu Thr Phe Pro Lys Asp
50 55 60
Tyr Pro Leu Arg Pro Pro Lys Met Lys Phe Ile Thr Glu Ile Trp His
65 70 75 80
Pro Asn Val Asp Lys Asn Gly Asp Val Cys Ile Ser Ile Leu His Glu
85 90 95
Pro Gly Glu Asp Lys Tyr Gly Tyr Glu Lys Pro Glu Glu Arg Trp Leu
100 105 110
Pro Ile His Thr Val Glu Thr Ile Met Ile Ser Val Ile Ser Met Leu
115 120 125
Ala Asp Pro Asn Gly Asp Ser Pro Ala Asn Val Asp Ala Ala Lys Glu
130 135 140
Trp Arg Glu Asp Arg Asn Gly Glu Phe Lys Arg Lys Val Ala Arg Cys

145 150 155 160
Val Arg Lys Ser Gln Glu Thr Ala Phe Glu
 165 170

[0334]

SEQ ID NO:23

SEQUENCE CHARACTERISTICS:

LENGTH: 510 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGACGGAGC TGCAGTCGGC ACTGCTACTG CGAAGACAGC TGGCAGAACT CAACAAAAAT 60
CCAGTGGGAAG GCTTTTCTGC AGGTTTAATA GATGACAATG ATCTCTACCG ATGGGAAGTC 120
CTTATTATTG GCCCTCCAGA TACACTTTAT GAAGGTGGTG TTTTAAAGGC TCATCTTACT 180
TTCCCAAAAG ATTATCCCCT CCGACCTCCT AAAATGAAAT TCATTACAGA AATCTGGCAC 240
CCAAATGTTG ATAAAAATGG TGATGTGTGC ATTTCTATTG TTCATGAGCC TGGGGAAGAT 300
AAGTATGGTT ATGAAAAGCC AGAGGAACGC TGGCTCCCTA TCCACACTGT GGAAACCATC 360
ATGATTAGTG TCATTTCTAT GCTGGCAGAC CCTAATGGAG ACTCACCTGC TAATGTTGAT 420
GCTGCCGAAAG AATGGAGGGA AGATAGAAAT GGAGAATTTA AAAGAAAAGT TGCCCGCTGT 480
GTAAGAAAAA GCCAAGAGAC TGCTTTTGAG 510

[0335]

SEQ ID NO:24

SEQUENCE CHARACTERISTICS:

LENGTH: 617 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-423A12

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 19..528

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

GGGCCCTCGG CAGGGAGG ATG ACG GAG CTG CAG TCG GCA CTG CTA CTG CGA	51
Met Thr Glu Leu Gln Ser Ala Leu Leu Leu Arg	
1 5 10	
AGA CAG CTG GCA GAA CTC AAC AAA AAT CCA GTG GAA GGC TTT TCT GCA	99
Arg Gln Leu Ala Glu Leu Asn Lys Asn Pro Val Glu Gly Phe Ser Ala	
15 20 25	
GGT TTA ATA GAT GAC AAT GAT CTC TAC CGA TGG GAA GTC CTT ATT ATT	147
Gly Leu Ile Asp Asp Asn Asp Leu Tyr Arg Trp Glu Val Leu Ile Ile	
30 35 40	
GGC CCT CCA GAT ACA CTT TAT GAA GGT GGT GTT TTT AAG GCT CAT CTT	195
Gly Pro Pro Asp Thr Leu Tyr Glu Gly Gly Val Phe Lys Ala His Leu	
45 50 55	
ACT TTC CCA AAA GAT TAT CCC CTC CGA CCT CCT AAA ATG AAA TTC ATT	243
Thr Phe Pro Lys Asp Tyr Pro Leu Arg Pro Pro Lys Met Lys Phe Ile	
60 65 70 75	
ACA GAA ATC TGG CAC CCA AAT GTT GAT AAA AAT GGT GAT GTG TGC ATT	291
Thr Glu Ile Trp His Pro Asn Val Asp Lys Asn Gly Asp Val Cys Ile	
80 85 90	

TCT ATT CTT CAT GAG CCT GGG GAA GAT AAG TAT GGT TAT GAA AAG CCA	339
Ser Ile Leu His Glu Pro Gly Glu Asp Lys Tyr Gly Tyr Glu Lys Pro	
95 100 105	
GAG GAA CGC TGG CTC CCT ATC CAC ACT GTG GAA ACC ATC ATG ATT AGT	387
Glu Glu Arg Trp Leu Pro Ile His Thr Val Glu Thr Ile Met Ile Ser	
110 115 120	
GTC ATT TCT ATG CTG GCA GAC CCT AAT GGA GAC TCA CCT GCT AAT GTT	435
Val Ile Ser Met Leu Ala Asp Pro Asn Gly Asp Ser Pro Ala Asn Val	
125 130 135	
GAT GCT GCG AAA GAA TGG AGG GAA GAT AGA AAT GGA GAA TTT AAA AGA	483
Asp Ala Ala Lys Glu Trp Arg Glu Asp Arg Asn Gly Glu Phe Lys Arg	
140 145 150 155	
AAA GTT GCC CGC TGT GTA AGA AAA AGC CAA GAG ACT GCT TTT GAG	528
Lys Val Ala Arg Cys Val Arg Lys Ser Gln Glu Thr Ala Phe Glu	
160 165 170	
TGACATTAT TTAGCAGCTA GTAAC TTCAC TTATTTTCAGG GTC TCCAATT GAGAAACATG	588
GCACTGTTTT TCCTGCACTC TACCCACCG	617

[0336]

SEQ ID NO:25

SEQUENCE CHARACTERISTICS:

LENGTH: 374 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met Val Leu Trp Glu Ser Pro Arg Gln Cys Ser Ser Trp Thr Leu Cys	
1 5 10 15	
Glu Gly Phe Cys Trp Leu Leu Leu Leu Pro Val Met Leu Leu Ile Val	
20 25 30	
Ala Arg Pro Val Lys Leu Ala Ala Phe Pro Thr Ser Leu Ser Asp Cys	
35 40 45	

Gln Thr Pro Thr Gly Trp Asn Cys Ser Gly Tyr Asp Asp Arg Glu Asn
 50 55 60
 Asp Leu Phe Leu Cys Asp Thr Asn Thr Cys Lys Phe Asp Gly Glu Cys
 65 70 75 80
 Leu Arg Ile Gly Asp Thr Val Thr Cys Val Cys Gln Phe Lys Cys Asn
 85 90 95
 Asn Asp Tyr Val Pro Val Cys Gly Ser Asn Gly Glu Ser Tyr Gln Asn
 100 105 110
 Glu Cys Tyr Leu Arg Gln Ala Ala Cys Lys Gln Gln Ser Glu Ile Leu
 115 120 125
 Val Val Ser Glu Gly Ser Cys Ala Thr Asp Ala Gly Ser Gly Ser Gly
 130 135 140
 Asp Gly Val His Glu Gly Ser Gly Glu Thr Ser Gln Lys Glu Thr Ser
 145 150 155 160
 Thr Cys Asp Ile Cys Gln Phe Gly Ala Glu Cys Asp Glu Asp Ala Glu
 165 170 175
 Asp Val Trp Cys Val Cys Asn Ile Asp Cys Ser Gln Thr Asn Phe Asn
 180 185 190
 Pro Leu Cys Ala Ser Asp Gly Lys Ser Tyr Asp Asn Ala Cys Gln Ile
 195 200 205
 Lys Glu Ala Ser Cys Gln Lys Gln Glu Lys Ile Glu Val Met Ser Leu
 210 215 220
 Gly Arg Cys Gln Asp Asn Thr Thr Thr Thr Thr Lys Ser Glu Asp Gly
 225 230 235 240
 His Tyr Ala Arg Thr Asp Tyr Ala Glu Asn Ala Asn Lys Leu Glu Glu
 245 250 255
 Ser Ala Arg Glu His His Ile Pro Cys Pro Glu His Tyr Asn Gly Phe
 260 265 270
 Cys Met His Gly Lys Cys Glu His Ser Ile Asn Met Gln Glu Pro Ser
 275 280 285
 Cys Arg Cys Asp Ala Gly Tyr Thr Gly Gln His Cys Glu Lys Lys Asp
 290 295 300
 Tyr Ser Val Leu Tyr Val Val Pro Gly Pro Val Arg Phe Gln Tyr Val

305		310		315		320									
Leu	Ile	Ala	Ala	Val	Ile	Gly	Thr	Ile	Gln	Ile	Ala	Val	Ile	Cys	Val
				325					330					335	
Val	Val	Leu	Cys	Ile	Thr	Arg	Lys	Cys	Pro	Arg	Ser	Asn	Arg	Ile	His
			340					345					350		
Arg	Gln	Lys	Gln	Asn	Thr	Gly	His	Tyr	Ser	Ser	Asp	Asn	Thr	Thr	Arg
		355					360					365			
Ala	Ser	Thr	Arg	Leu	Ile										
		370													

[0337]

SEQ ID NO:26

SEQUENCE CHARACTERISTICS:

LENGTH: 1122 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGGTGCTGT	GGGAGTCCCC	GCGGCAGTGC	AGCAGCTGGA	CACTTTGCGA	GGGCTTTTGC	60
TGGCTGCTGC	TGCTGCCCCG	CATGCTACTC	ATCGTAGCCC	GCCCGGTGAA	GCTCGCTGCT	120
TTCCCTACCT	CCTTAAGTGA	CTGCCAAACG	CCCACCGGCT	GGAATTGCTC	TGGTTATGAT	180
GACAGAGAAA	ATGATCTCTT	CCTCTGTGAC	ACCAACACCT	GTAAATTTGA	TGGGGAATGT	240
TTAAGAATTG	GAGACACTGT	GACTTGCGTC	TGTCAGTTCA	AGTGCAACAA	TGACTATGTG	300
CCTGTGTGTG	GCTCCAATGG	GGAGAGCTAC	CAGAATGAGT	GTTACCTGCG	ACAGGCTGCA	360
TGCAAACAGC	AGAGTGAGAT	ACTTGTGGTG	TCAGAAGGAT	CATGTGCCAC	AGATGCAGGA	420
TCAGGATCTG	GAGATGGAGT	CCATGAAGGC	TCTGGAGAAA	CTAGTCAAAA	GGAGACATCC	480

ACCTGTGATA TTTGCCAGTT TGGTGCAGAA TGTGACGAAG ATGCCGAGGA TGTC'TGGTGT	540
GTGTGTAATA TTGACTGTTT TCAAACCAAC TTCAATCCCC TCTGCGCTTC TGATGGGAAA	600
TCTTATGATA ATGCATGCCA AATCAAAGAA GCATCGTGTC AGAAACAGGA GAAAATTGAA	660
GTCATGTC'TT TGGGTCGATG TCAAGATAAC ACAACTACAA CTACTAAGTC TGAAGATGGG	720
CATTATGCAA GAACAGATTA TGCAGAGAAT GCTAACAAAT TAGAAGAAAG TGCCAGAGAA	780
CACCACATAC CTTGTCCGGA ACATTACAAT GGCTTCTGCA TGCATGGGAA GTGTGAGCAT	840
TCTATCAATA TGCAGGAGCC ATCTTGCAGG TGTGATGCTG GTTATACTGG ACAACACTGT	900
GAAAAAAAGG ACTACAGTGT TCTATACGTT GTTCCCGGTC CTGTACGATT TCAGTATGTC	960
TTAATCGCAG CTGTGATTGG AACAAATCAG ATTGCTGTCA TCTGTGTGGT GGTCCCTCTGC	1020
ATCACAAGGA AATGCCCCAG AAGCAACAGA ATTCACAGAC AGAAGCAAAA TACAGGGCAC	1080
TACAGTTCAG ACAATACAAC AAGAGCGTCC ACGAGGTTAA TC	1122

[0338]

SEQ ID NO:27

SEQUENCE CHARACTERISTICS:

LENGTH: 1721 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-092E10

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 368..1489

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

CTGCGGGGCG CCTTGACTCT CCCGCCACCC TGCTCTCTCG GGCTCCACTC GTCTGCCCCCT	60
GGACTCCCGT CTCCTCCTGT CCTCCGGCTT CCCAGAGCTC CCTCCTTATG GCAGCAGCTT	120
CCCGCGTCTC CGGCGCAGCT TCTCAGCGGA CGACCTCTC GCTCCGGGGC TGAGCCAGTC	180
CCTGGATGTT GCTGAAACTC TCGAGATCAT GCGCGGGTTT GGCTGCTGCT TCCCCGCCGG	240
GTGCCACTGC CACCGCCGCC GCCTCTGCTG CCGCCGTCCG CGGGATGCTC AGTAGCCCGC	300
TGCCCCGCCC CCGCGATCCT GTGTTCTCTG GAAGCCGTTT GCTGCTGCAG AGTTGCACGA	360
ACTAGTC ATG GTG CTG TGG GAG TCC CCG CGG CAG TGC AGC AGC TGG ACA	409
Met Val Leu Trp Glu Ser Pro Arg Gln Cys Ser Ser Trp Thr	
1 5 10	
CTT TGC GAG GGC TTT TGC TGG CTG CTG CTG CTG CCC GTC ATG CTA CTC	457
Leu Cys Glu Gly Phe Cys Trp Leu Leu Leu Leu Pro Val Met Leu Leu	
15 20 25 30	
ATC GTA GCC CGC CCG GTG AAG CTC GCT GCT TTC CCT ACC TCC TTA AGT	505
Ile Val Ala Arg Pro Val Lys Leu Ala Ala Phe Pro Thr Ser Leu Ser	
35 40 45	
GAC TGC CAA ACG CCC ACC GGC TGG AAT TGC TCT GGT TAT GAT GAC AGA	553
Asp Cys Gln Thr Pro Thr Gly Trp Asn Cys Ser Gly Tyr Asp Asp Arg	
50 55 60	
GAA AAT GAT CTC TTC CTC TGT GAC ACC AAC ACC TGT AAA TTT GAT GGG	601
Glu Asn Asp Leu Phe Leu Cys Asp Thr Asn Thr Cys Lys Phe Asp Gly	
65 70 75	
GAA TGT TTA AGA ATT GGA GAC ACT GTG ACT TGC GTC TGT CAG TTC AAG	649
Glu Cys Leu Arg Ile Gly Asp Thr Val Thr Cys Val Cys Gln Phe Lys	
80 85 90	
TGC AAC AAT GAC TAT GTG CCT GTG TGT GGC TCC AAT GGG GAG AGC TAC	697
Cys Asn Asn Asp Tyr Val Pro Val Cys Gly Ser Asn Gly Glu Ser Tyr	
95 100 105 110	
CAG AAT GAG TGT TAC CTG CGA CAG GCT GCA TGC AAA CAG CAG AGT GAG	745
Gln Asn Glu Cys Tyr Leu Arg Gln Ala Ala Cys Lys Gln Gln Ser Glu	
115 120 125	

ATA CTT GTG GTG TCA GAA GGA TCA TGT GCC ACA GAT GCA GGA TCA GGA	793
Ile Leu Val Val Ser Glu Gly Ser Cys Ala Thr Asp Ala Gly Ser Gly	
130 135 140	
TCT GGA GAT GGA GTC CAT GAA GGC TCT GGA GAA ACT AGT CAA AAG GAG	841
Ser Gly Asp Gly Val His Glu Gly Ser Gly Glu Thr Ser Gln Lys Glu	
145 150 155	
ACA TCC ACC TGT GAT ATT TGC CAG TTT GGT GCA GAA TGT GAC GAA GAT	889
Thr Ser Thr Cys Asp Ile Cys Gln Phe Gly Ala Glu Cys Asp Glu Asp	
160 165 170	
GCC GAG GAT GTC TGG TGT GTG TGT AAT ATT GAC TGT TCT CAA ACC AAC	937
Ala Glu Asp Val Trp Cys Val Cys Asn Ile Asp Cys Ser Gln Thr Asn	
175 180 185 190	
TTC AAT CCC CTC TGC GCT TCT GAT GGG AAA TCT TAT GAT AAT GCA TGC	985
Phe Asn Pro Leu Cys Ala Ser Asp Gly Lys Ser Tyr Asp Asn Ala Cys	
195 200 205	
CAA ATC AAA GAA GCA TCG TGT CAG AAA CAG GAG AAA ATT GAA GTC ATG	1033
Gln Ile Lys Glu Ala Ser Cys Gln Lys Gln Glu Lys Ile Glu Val Met	
210 215 220	
TCT TTG GGT CGA TGT CAA GAT AAC ACA ACT ACA ACT ACT AAG TCT GAA	1081
Ser Leu Gly Arg Cys Gln Asp Asn Thr Thr Thr Thr Thr Lys Ser Glu	
225 230 235	
GAT GGG CAT TAT GCA AGA ACA GAT TAT GCA GAG AAT GCT AAC AAA TTA	1129
Asp Gly His Tyr Ala Arg Thr Asp Tyr Ala Glu Asn Ala Asn Lys Leu	
240 245 250	
GAA GAA AGT GCC AGA GAA CAC CAC ATA CCT TGT CCG GAA CAT TAC AAT	1177
Glu Glu Ser Ala Arg Glu His His Ile Pro Cys Pro Glu His Tyr Asn	
255 260 265 270	
GGC TTC TGC ATG CAT GGG AAG TGT GAG CAT TCT ATC AAT ATG CAG GAG	1225
Gly Phe Cys Met His Gly Lys Cys Glu His Ser Ile Asn Met Gln Glu	
275 280 285	
CCA TCT TGC AGG TGT GAT GCT GGT TAT ACT GGA CAA CAC TGT GAA AAA	1273
Pro Ser Cys Arg Cys Asp Ala Gly Tyr Thr Gly Gln His Cys Glu Lys	
290 295 300	
AAG GAC TAC AGT GTT CTA TAC GTT GTT CCC GGT CCT GTA CGA TTT CAG	1321
Lys Asp Tyr Ser Val Leu Tyr Val Val Pro Gly Pro Val Arg Phe Gln	
305 310 315	
TAT GTC TTA ATC GCA GCT GTG ATT GGA ACA ATT CAG ATT GCT GTC ATC	1369

-153-

Tyr	Val	Leu	Ile	Ala	Ala	Val	Ile	Gly	Thr	Ile	Gln	Ile	Ala	Val	Ile		
320						325					330						
TGT	GTG	GTG	GTC	CTC	TGC	ATC	ACA	AGG	AAA	TGC	CCC	AGA	AGC	AAC	AGA	1417	
Cys	Val	Val	Val	Leu	Cys	Ile	Thr	Arg	Lys	Cys	Pro	Arg	Ser	Asn	Arg		
335					340					345					350		
ATT	CAC	AGA	CAG	AAG	CAA	AAT	ACA	GGG	CAC	TAC	AGT	TCA	GAC	AAT	ACA	1465	
Ile	His	Arg	Gln	Lys	Gln	Asn	Thr	Gly	His	Tyr	Ser	Ser	Asp	Asn	Thr		
				355					360						365		
ACA	AGA	GCG	TCC	ACG	AGG	TTA	ATC	TAA	AGGGAGCATG	TTTCACAGTG						1512	
Thr	Arg	Ala	Ser	Thr	Arg	Leu	Ile										
				370													
GCTGGACTAC	CGAGAGCTTG	GACTACACAA	TACAGTATTA	TAGACAAAAG	AATAAGACAA											1572	
GAGATCTACA	CATGTTGCCT	TGCATTTGTG	GTAATCTACA	CCAATGAAAA	CATGTACTAC											1632	
AGCTATATTT	GATTATGTAT	GGATATATTT	GAAATAGTAT	ACATTGTCTT	GATGTTTTTT											1692	
CTGTAATGTA	AATAAACTAT	TTATATCAC														1721	

[0339]

SEQ ID NO:28

SEQUENCE CHARACTERISTICS:

LENGTH: 817 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met	Gly	Asp	Thr	Val	Val	Glu	Pro	Ala	Pro	Leu	Lys	Pro	Thr	Ser	Glu
1				5					10					15	
Pro	Thr	Ser	Gly	Pro	Pro	Gly	Asn	Asn	Gly	Gly	Ser	Leu	Leu	Ser	Val
			20				25						30		
Ile	Thr	Glu	Gly	Val	Gly	Glu	Leu	Ser	Val	Ile	Asp	Pro	Glu	Val	Ala
		35					40					45			

-154-

Gln Lys Ala Cys Gln Glu Val Leu Glu Lys Val Lys Leu Leu His Gly
50 55 60

Gly Val Ala Val Ser Ser Arg Gly Thr Pro Leu Glu Leu Val Asn Gly
65 70 75 80

Asp Gly Val Asp Ser Glu Ile Arg Cys Leu Asp Asp Pro Pro Ala Gln
85 90 95

Ile Arg Glu Glu Glu Asp Glu Met Gly Ala Ala Val Ala Ser Gly Thr
100 105 110

Ala Lys Gly Ala Arg Arg Arg Arg Gln Asn Asn Ser Ala Lys Gln Ser
115 120 125

Trp Leu Leu Arg Leu Phe Glu Ser Lys Leu Phe Asp Ile Ser Met Ala
130 135 140

Ile Ser Tyr Leu Tyr Asn Ser Lys Glu Pro Gly Val Gln Ala Tyr Ile
145 150 155 160

Gly Asn Arg Leu Phe Cys Phe Arg Asn Glu Asp Val Asp Phe Tyr Leu
165 170 175

Pro Gln Leu Leu Asn Met Tyr Ile His Met Asp Glu Asp Val Gly Asp
180 185 190

Ala Ile Lys Pro Tyr Ile Val His Arg Cys Arg Gln Ser Ile Asn Phe
195 200 205

Ser Leu Gln Cys Ala Leu Leu Leu Gly Ala Tyr Ser Ser Asp Met His
210 215 220

Ile Ser Thr Gln Arg His Ser Arg Gly Thr Lys Leu Arg Lys Leu Ile
225 230 235 240

Leu Ser Asp Glu Leu Lys Pro Ala His Arg Lys Arg Glu Leu Pro Ser
245 250 255

Leu Ser Pro Ala Pro Asp Thr Gly Leu Ser Pro Ser Lys Arg Thr His
260 265 270

Gln Arg Ser Lys Ser Asp Ala Thr Ala Ser Ile Ser Leu Ser Ser Asn
275 280 285

Leu Lys Arg Thr Ala Ser Asn Pro Lys Val Glu Asn Glu Asp Glu Glu
290 295 300

Leu Ser Ser Ser Thr Glu Ser Ile Asp Asn Ser Phe Ser Ser Pro Val

305		310		315		320
Arg Leu Ala Pro	Glu Arg Glu Phe Ile	Lys Ser Leu Met Ala	Ile Gly			
	325	330	335			
Lys Arg Leu Ala Thr	Leu Pro Thr Lys Glu Gln Lys Thr	Gln Arg Leu				
	340	345	350			
Ile Ser Glu Leu Ser	Leu Leu Asn His Lys Leu Pro Ala	Arg Val Trp				
	355	360	365			
Leu Pro Thr Ala Gly	Phe Asp His His Val Val Arg Val	Pro His Thr				
	370	375	380			
Gln Ala Val Val Leu	Asn Ser Lys Asp Lys Ala Pro Tyr Leu	Ile Tyr				
	385	390	395			400
Val Glu Val Leu Glu	Cys Glu Asn Phe Asp Thr Thr Ser Val	Pro Ala				
	405	410	415			
Arg Ile Pro Glu Asn	Arg Ile Arg Ser Thr Arg Ser Val	Glu Asn Leu				
	420	425	430			
Pro Glu Cys Gly Ile	Thr His Glu Gln Arg Ala Gly Ser Phe	Ser Thr				
	435	440	445			
Val Pro Asn Tyr Asp	Asn Asp Asp Glu Ala Trp Ser Val Asp Asp	Ile				
	450	455	460			
Gly Glu Leu Gln Val	Glu Leu Pro Glu Val His Thr Asn Ser Cys	Asp				
	465	470	475			480
Asn Ile Ser Gln Phe	Ser Val Asp Ser Ile Thr Ser Gln Glu Ser	Lys				
	485	490	495			
Glu Pro Val Phe Ile	Ala Ala Gly Asp Ile Arg Arg Arg Leu Ser	Glu				
	500	505	510			
Gln Leu Ala His Thr	Pro Thr Ala Phe Lys Arg Asp Pro Glu Asp	Pro				
	515	520	525			
Ser Ala Val Ala Leu	Lys Glu Pro Trp Gln Glu Lys Val Arg Arg	Ile				
	530	535	540			
Arg Glu Gly Ser Pro	Tyr Gly His Leu Pro Asn Trp Arg Leu Leu	Ser				
	545	550	555			560
Val Ile Val Lys Cys	Gly Asp Asp Leu Arg Gln Glu Leu Leu Ala	Phe				
	565	570	575			

Gln	Val	Leu	Lys	Gln	Leu	Gln	Ser	Ile	Trp	Glu	Gln	Glu	Arg	Val	Pro
			580					585					590		
Leu	Trp	Ile	Lys	Pro	Ile	Gln	Asp	Ser	Cys	Glu	Ile	Thr	Thr	Asp	Ser
			595				600					605			
Gly	Met	Ile	Glu	Pro	Val	Val	Asn	Ala	Val	Ser	Ile	His	Gln	Val	Lys
	610					615					620				
Lys	Gln	Ser	Gln	Leu	Ser	Leu	Leu	Asp	Tyr	Phe	Leu	Gln	Glu	His	Gly
625					630					635					640
Ser	Tyr	Thr	Thr	Glu	Ala	Phe	Leu	Ser	Ala	Gln	Arg	Asn	Phe	Val	Gln
				645					650					655	
Ser	Cys	Ala	Gly	Tyr	Cys	Leu	Val	Cys	Tyr	Leu	Leu	Gln	Val	Lys	Asp
			660					665					670		
Arg	His	Asn	Gly	Asn	Ile	Leu	Leu	Asp	Ala	Glu	Gly	His	Ile	Ile	His
	675					680						685			
Ile	Asp	Phe	Gly	Phe	Ile	Leu	Ser	Ser	Ser	Pro	Arg	Asn	Leu	Gly	Phe
	690					695					700				
Glu	Thr	Ser	Ala	Phe	Lys	Leu	Thr	Thr	Glu	Phe	Val	Asp	Val	Met	Gly
705					710					715					720
Gly	Leu	Asp	Gly	Asp	Met	Phe	Asn	Tyr	Tyr	Lys	Met	Leu	Met	Leu	Gln
				725					730					735	
Gly	Leu	Ile	Ala	Ala	Arg	Lys	His	Met	Asp	Lys	Val	Val	Gln	Ile	Val
			740					745					750		
Glu	Ile	Met	Gln	Gln	Gly	Ser	Gln	Leu	Pro	Cys	Phe	His	Gly	Ser	Ser
		755					760					765			
Thr	Ile	Arg	Asn	Leu	Lys	Glu	Arg	Phe	His	Met	Ser	Met	Thr	Glu	Glu
	770					775					780				
Gln	Leu	Gln	Leu	Leu	Val	Glu	Gln	Met	Val	Asp	Gly	Ser	Met	Arg	Ser
785				790						795					800
Ile	Thr	Thr	Lys	Leu	Tyr	Asp	Gly	Phe	Gln	Tyr	Leu	Thr	Asn	Gly	Ile
			805						810					815	
Met															

SEQ ID NO:29

SEQUENCE CHARACTERISTICS:

LENGTH: 2451 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGGGAGATA CAGTAGTGGA GCCTGCCCCC TTGAAGCCAA CTTCTGAGCC CACTTCTGGC	60
CCACCAGGGA ATAATGGGGG GTCCCTGCTA AGTGTCATCA CGGAGGGGGT CGGGGAACTA	120
TCAGTGATG ACCCTGAGGT GGCCCAGAAG GCCTGCCAGG AGGTGTTGGA GAAAGTCAAG	180
CTTTTGCATG GAGGCGTGGC AGTCTCTAGC AGAGGCACCC CACTGGAGTT GGTCAATGGG	240
GATGGTGIGG ACAGTGAGAT CCGTTGCCTA GATGATCCAC CTGCCCAGAT CAGGGAGGAG	300
GAAGATGAGA TGGGGGCCGC TGTGGCCTCA GGCACAGCCA AAGGAGCAAG AAGACGGCGG	360
CAGAACAACT CAGCTAAACA GTCTTGGCTG CTGAGGCTGT TTGAGTCAAA ACTGTTTGAC	420
ATCTCCATGG CCATTTTATA CCTGTATAAC TCCAAGGAGC CTGGAGTACA AGCCTACATT	480
GGCAACCGGC TCTTCTGCTT TCGCAACGAG GACGTGGACT TCTATCTGCC CCAGTTGCTT	540
AACATGTACA TCCACATGGA TGAGGACGTG GGTGATGCCA TTAAGCCCTA CATAGTCCAC	600
CGTTGCCGCC AGAGCATTAA CTTTTCCTC CAGTGTGCCC TGTGCTTGG GGCCTATTCT	660
TCAGACATGC ACATTTCCAC TCAACGACAC TCCCGTGGGA CCAAGCTACG GAAGCTGATC	720
CTCTCAGATG AGCTAAAGCC AGCTCACAGG AAGAGGGAGC TGCCCTCCTT GAGCCCGGCC	780
CCTGATACAG GGCTGTCTCC CTCCAAAAGG ACTCACCAGC GCTCTAAGTC AGATGCCACT	840
GCCAGCATAA GTCTCAGCAG CAACCTGAAA CGAACAGCCA GCAACCCCTAA AGTGGAGAAT	900
GAGGATGAGG AGCTCTCCTC CAGCACCGAG AGTATTGATA ATTCAATCAG TTCCCCTGTT	960
CGACTGGCTC CTGAGAGAGA ATTCATCAAG TCCCTGATGG CGATCGGCAA GCGGCTGGCC	1020

ACGCTCCCCA CCAAAGAGCA GAAAACACAG AGGCTGATCT CAGAGCTCTC CCTGCTCAAC	1080
CATAAGCTCC CTGCCCCGAGT CTGGCTGCCC ACTGCTGGCT TTGACCACCA CGTGGTCCGT	1140
GTACCCCAACA CACAGGCTGT TGTCTCAAC TCCAAGGACA AGGCTCCCTA CCTGATTTAT	1200
GTGGAAGTCC TTGAATGTGA AAACCTTTGAC ACCACCAGTG TCCCTGCCCCG GATCCCCGAG	1260
AACCGAATTC GGAGTACGAG GTCCGTAGAA AACTTGCCCCG AATGTGGTAT TACCCATGAG	1320
CAGCGAGCTG GCAGCTTCAG CACTGTGCCC AACTATGACA ACGATGATGA GGCCTGGTCG	1380
GTGGATGACA TAGGCGAGCT GCAAGTGGAG CTCCCCGAAG TGCATACCAA CAGCTGTGAC	1440
AACATCTCCC AGTTCTCTGT GGACAGCATC ACCAGCCAGG AGAGCAAGGA GCCTGTGTTC	1500
ATTGCAGCAG GGGACATCCG CCGGCGCCTT TCGGAACAGC TGGCTCATAAC CCCGACAGCC	1560
TTCAAACGAG ACCCAGAAGA TCCTTCTGCA GTTGCTCTCA AAGAGCCCTG GCAGGAGAAA	1620
GTACGGCGGA TCAGAGAGGG CTCCCCCTAC GGCCATCTCC CCAATTGGCG GCTCCTGTCA	1680
GTCAATTGTCA AGTGTGGGGA TGACCTTCGG CAAGAGCTTC TGGCCTTTCA GGTGTGAAG	1740
CAACTGCAGT CCAATTGGGA ACAGGAGCGA GTGCCCCCTT GGATCAAGCC AATACAAGAT	1800
TCTTGTGAAA TTACGACTGA TAGTGGCATG ATTGAACCAG TGGTCAATGC TGTGTCCATC	1860
CATCAGGTGA AGAAACAGTC ACAGCTCTCC TTGCTCGATT ACTTCCTACA GGAGCACGGC	1920
AGTTACACCA CTGAGGCATT CCTCAGTGCA CAGCGCAATT TTGTGCAAAG TTGTGCTGGG	1980
TACTGCTTGG TCTGCTACCT GCTGCAAGTC AAGGACAGAC ACAATGGGAA TATCCTTTTG	2040
GACGCAGAAG GCCACATCAT CCACATCGAC TTTGGCTTCA TCCTCTCCAG CTCACCCCGA	2100
AATCTGGGCT TTGAGACGTC AGCCTTTAAG CTGACCACAG AGTTTGIGGA TGTGATGGGC	2160
GGCCTGGATG GCGACATGTT CAACTACTAT AAGATGCTGA TGCTGCAAGG GCTGATTGCC	2220
GCTCGGAAAC ACATGGACAA GGTGGTGCAG ATCGTGGAGA TCATGCAGCA AGGTTCTCAG	2280
CTTCCTTGCT TCCATGGCTC CAGCACCATT CGAAACCTCA AAGAGAGGTT CCACATGAGC	2340
ATGACTGAGG AGCAGCTGCA GCTGCTGGTG GAGCAGATGG TGGATGGCAG TATGCGGTCT	2400
ATCACCACCA AACTCTATGA CGGCTTCCAG TACCTCACCA ACGGCATCAT G	2451

[0341]

SEQ ID NO:30

SEQUENCE CHARACTERISTICS:

LENGTH: 3602 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-428B12c2

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 429..2879

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

GGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGGACAA GGCAGATCCC TTGAGCCCAG	60
GAGGTAGAGG CTCCAGTGAG CTGTGATGGT GCCACTGCAC TCCAGCCTGG GCAATGAAGC	120
AAGACCCTAT CTGAAAAAAAA AAATTTTAA AAAAGGCAAA GATGGGCCTG GGGCACCAAA	180
TATTCCAGAG GAAAGGGAAC GTGTGTACTC CTTGAGGTGG GGAACATGAC CCACTTGAGG	240
TGCAGAAAGA AGACTTGTAT GGGGCTGGTG CAGCCTCCGC GGCCGCTGTC AGGGAAGCGC	300
AGGCGGCCAA TGGAACCCGG GAGCGGTCGC TGCTGCTGAG GCGGCAGTGT CGGCAGTCCA	360
ACCGCGACTG CCCGCACCCC CTCCGCGGGG TCCCCCAGAG CTTGGAAGCT CGAAGTCTGG	420
CTGTGGCC ATG GGA GAT ACA GTA GTG GAG CCT GCC CCC TTG AAG CCA ACT	470
Met Gly Asp Thr Val Val Glu Pro Ala Pro Leu Lys Pro Thr	
1 5 10	

TCT	GAG	CCC	ACT	TCT	GGC	CCA	CCA	GGG	AAT	AAT	GGG	GGG	TCC	CTG	CTA	518
Ser	Glu	Pro	Thr	Ser	Gly	Pro	Pro	Gly	Asn	Asn	Gly	Gly	Ser	Leu	Leu	
15					20				25						30	
AGT	GTC	ATC	ACG	GAG	GGG	GTC	GGG	GAA	CTA	TCA	GTG	ATT	GAC	CCT	GAG	566
Ser	Val	Ile	Thr	Glu	Gly	Val	Gly	Glu	Leu	Ser	Val	Ile	Asp	Pro	Glu	
				35				40						45		
GTG	GCC	CAG	AAG	GCC	TGC	CAG	GAG	GTG	TTG	GAG	AAA	GTC	AAG	CTT	TTG	614
Val	Ala	Gln	Lys	Ala	Cys	Gln	Glu	Val	Leu	Glu	Lys	Val	Lys	Leu	Leu	
			50					55						60		
CAT	GGA	GGC	GTG	GCA	GTC	TCT	AGC	AGA	GGC	ACC	CCA	CTG	GAG	TTG	GTC	662
His	Gly	Gly	Val	Ala	Val	Ser	Ser	Arg	Gly	Thr	Pro	Leu	Glu	Leu	Val	
		65					70						75			
AAT	GGG	GAT	GGT	GTG	GAC	AGT	GAG	ATC	CGT	TGC	CTA	GAT	GAT	CCA	CCT	710
Asn	Gly	Asp	Gly	Val	Asp	Ser	Glu	Ile	Arg	Cys	Leu	Asp	Asp	Pro	Pro	
	80					85					90					
GCC	CAG	ATC	AGG	GAG	GAG	GAA	GAT	GAG	ATG	GGG	GCC	GCT	GTG	GCC	TCA	758
Ala	Gln	Ile	Arg	Glu	Glu	Glu	Asp	Glu	Met	Gly	Ala	Ala	Val	Ala	Ser	
95					100					105					110	
GGC	ACA	GCC	AAA	GGA	GCA	AGA	AGA	CGG	CGG	CAG	AAC	AAC	TCA	GCT	AAA	806
Gly	Thr	Ala	Lys	Gly	Ala	Arg	Arg	Arg	Arg	Gln	Asn	Asn	Ser	Ala	Lys	
				115				120						125		
CAG	TCT	TGG	CTG	CTG	AGG	CTG	TTT	GAG	TCA	AAA	CTG	TTT	GAC	ATC	TCC	854
Gln	Ser	Trp	Leu	Leu	Arg	Leu	Phe	Glu	Ser	Lys	Leu	Phe	Asp	Ile	Ser	
			130					135					140			
ATG	GCC	ATT	TCA	TAC	CTG	TAT	AAC	TCC	AAG	GAG	CCT	GGA	GTA	CAA	GCC	902
Met	Ala	Ile	Ser	Tyr	Leu	Tyr	Asn	Ser	Lys	Glu	Pro	Gly	Val	Gln	Ala	
		145					150					155				
TAC	ATT	GGC	AAC	CGG	CTC	TTC	TGC	TTT	CGC	AAC	GAG	GAC	GTG	GAC	TTC	950
Tyr	Ile	Gly	Asn	Arg	Leu	Phe	Cys	Phe	Arg	Asn	Glu	Asp	Val	Asp	Phe	
	160					165					170					
TAT	CTG	CCC	CAG	TTG	CTT	AAC	ATG	TAC	ATC	CAC	ATG	GAT	GAG	GAC	GTG	998
Tyr	Leu	Pro	Gln	Leu	Leu	Asn	Met	Tyr	Ile	His	Met	Asp	Glu	Asp	Val	
175					180					185					190	
GGT	GAT	GCC	ATT	AAG	CCC	TAC	ATA	GTC	CAC	CGT	TGC	CGC	CAG	AGC	ATT	1046
Gly	Asp	Ala	Ile	Lys	Pro	Tyr	Ile	Val	His	Arg	Cys	Arg	Gln	Ser	Ile	
				195					200					205		
AAC	TTT	TCC	CTC	CAG	TGT	GCC	CTG	TTG	CTT	GGG	GCC	TAT	TCT	TCA	GAC	1094

Asn	Phe	Ser	Leu	Gln	Cys	Ala	Leu	Leu	Leu	Gly	Ala	Tyr	Ser	Ser	Asp	
			210					215					220			
ATG	CAC	ATT	TCC	ACT	CAA	CGA	CAC	TCC	CGT	GGG	ACC	AAG	CTA	CGG	AAG	1142
Met	His	Ile	Ser	Thr	Gln	Arg	His	Ser	Arg	Gly	Thr	Lys	Leu	Arg	Lys	
		225					230					235				
CTG	ATC	CTC	TCA	GAT	GAG	CTA	AAG	CCA	GCT	CAC	AGG	AAG	AGG	GAG	CTG	1190
Leu	Ile	Leu	Ser	Asp	Glu	Leu	Lys	Pro	Ala	His	Arg	Lys	Arg	Glu	Leu	
	240					245					250					
CCC	TCC	TTG	AGC	CCG	GCC	CCT	GAT	ACA	GGG	CTG	TCT	CCC	TCC	AAA	AGG	1238
Pro	Ser	Leu	Ser	Pro	Ala	Pro	Asp	Thr	Gly	Leu	Ser	Pro	Ser	Lys	Arg	
255					260				265						270	
ACT	CAC	CAG	CGC	TCT	AAG	TCA	GAT	GCC	ACT	GCC	AGC	ATA	AGT	CTC	AGC	1286
Thr	His	Gln	Arg	Ser	Lys	Ser	Asp	Ala	Thr	Ala	Ser	Ile	Ser	Leu	Ser	
				275					280					285		
AGC	AAC	CTG	AAA	CGA	ACA	GCC	AGC	AAC	CCT	AAA	GIG	GAG	AAT	GAG	GAT	1334
Ser	Asn	Leu	Lys	Arg	Thr	Ala	Ser	Asn	Pro	Lys	Val	Glu	Asn	Glu	Asp	
		290						295					300			
GAG	GAG	CTC	TCC	TCC	AGC	ACC	GAG	AGT	ATT	GAT	AAT	TCA	TTC	AGT	TCC	1382
Glu	Glu	Leu	Ser	Ser	Ser	Thr	Glu	Ser	Ile	Asp	Asn	Ser	Phe	Ser	Ser	
		305					310					315				
CCT	GTT	CGA	CTG	GCT	CCT	GAG	AGA	GAA	TTC	ATC	AAG	TCC	CTG	ATG	GCG	1430
Pro	Val	Arg	Leu	Ala	Pro	Glu	Arg	Glu	Phe	Ile	Lys	Ser	Leu	Met	Ala	
	320					325					330					
ATC	GGC	AAG	CGG	CTG	GCC	ACG	CTC	CCC	ACC	AAA	GAG	CAG	AAA	ACA	CAG	1478
Ile	Gly	Lys	Arg	Leu	Ala	Thr	Leu	Pro	Thr	Lys	Glu	Gln	Lys	Thr	Gln	
335					340					345					350	
AGG	CTG	ATC	TCA	GAG	CTC	TCC	CTG	CTC	AAC	CAT	AAG	CTC	CCT	GCC	CGA	1526
Arg	Leu	Ile	Ser	Glu	Leu	Ser	Leu	Leu	Asn	His	Lys	Leu	Pro	Ala	Arg	
				355					360					365		
GTC	TGG	CTG	CCC	ACT	GCT	GGC	TTT	GAC	CAC	CAC	GIG	GTC	CGT	GTA	CCC	1574
Val	Trp	Leu	Pro	Thr	Ala	Gly	Phe	Asp	His	His	Val	Val	Arg	Val	Pro	
		370					375					380				
CAC	ACA	CAG	GCT	GTT	GTC	CTC	AAC	TCC	AAG	GAC	AAG	GCT	CCC	TAC	CTG	1622
His	Thr	Gln	Ala	Val	Val	Leu	Asn	Ser	Lys	Asp	Lys	Ala	Pro	Tyr	Leu	
		385					390					395				
ATT	TAT	GTG	GAA	GTC	CTT	GAA	TGT	GAA	AAC	TTT	GAC	ACC	ACC	AGT	GTC	1670
Ile	Tyr	Val	Glu	Val	Leu	Glu	Cys	Glu	Asn	Phe	Asp	Thr	Thr	Ser	Val	

400	405	410	
CCT GCC CGG ATC CCC GAG AAC CGA ATT CGG AGT ACG AGG TCC GTA GAA Pro Ala Arg Ile Pro Glu Asn Arg Ile Arg Ser Thr Arg Ser Val Glu 415 420 425 430			1718
AAC TTG CCC GAA TGT GGT ATT ACC CAT GAG CAG CGA GCT GGC AGC TTC Asn Leu Pro Glu Cys Gly Ile Thr His Glu Gln Arg Ala Gly Ser Phe 435 440 445			1766
AGC ACT GTG CCC AAC TAT GAC AAC GAT GAT GAG GCC TGG TCG GTG GAT Ser Thr Val Pro Asn Tyr Asp Asn Asp Asp Glu Ala Trp Ser Val Asp 450 455 460			1814
GAC ATA GGC GAG CTG CAA GTG GAG CTC CCC GAA GTG CAT ACC AAC AGC Asp Ile Gly Glu Leu Gln Val Glu Leu Pro Glu Val His Thr Asn Ser 465 470 475			1862
TGT GAC AAC ATC TCC CAG TTC TCT GTG GAC AGC ATC ACC AGC CAG GAG Cys Asp Asn Ile Ser Gln Phe Ser Val Asp Ser Ile Thr Ser Gln Glu 480 485 490			1910
AGC AAG GAG CCT GTG TTC ATT GCA GCA GGG GAC ATC CGC CGG CGC CTT Ser Lys Glu Pro Val Phe Ile Ala Ala Gly Asp Ile Arg Arg Arg Leu 495 500 505 510			1958
TCG GAA CAG CTG GCT CAT ACC CCG ACA GCC TTC AAA CGA GAC CCA GAA Ser Glu Gln Leu Ala His Thr Pro Thr Ala Phe Lys Arg Asp Pro Glu 515 520 525			2006
GAT CCT TCT GCA GTT GCT CTC AAA GAG CCC TGG CAG GAG AAA GTA CGG Asp Pro Ser Ala Val Ala Leu Lys Glu Pro Trp Gln Glu Lys Val Arg 530 535 540			2054
CGG ATC AGA GAG GGC TCC CCC TAC GGC CAT CTC CCC AAT TGG CGG CTC Arg Ile Arg Glu Gly Ser Pro Tyr Gly His Leu Pro Asn Trp Arg Leu 545 550 555			2102
CTG TCA GTC ATT GTC AAG TGT GGG GAT GAC CTT CGG CAA GAG CTT CTG Leu Ser Val Ile Val Lys Cys Gly Asp Asp Leu Arg Gln Glu Leu Leu 560 565 570			2150
GCC TTT CAG GTG TTG AAG CAA CTG CAG TCC ATT TGG GAA CAG GAG CGA Ala Phe Gln Val Leu Lys Gln Leu Gln Ser Ile Trp Glu Gln Glu Arg 575 580 585 590			2198
GTG CCC CTT TGG ATC AAG CCA ATA CAA GAT TCT TGT GAA ATT ACG ACT Val Pro Leu Trp Ile Lys Pro Ile Gln Asp Ser Cys Glu Ile Thr Thr 595 600 605			2246

GAT AGT GGC ATG ATT GAA CCA GTG GTC AAT GCT GTG TCC ATC CAT CAG Asp Ser Gly Met Ile Glu Pro Val Val Asn Ala Val Ser Ile His Gln 610 615 620	2294
GTG AAG AAA CAG TCA CAG CTC TCC TTG CTC GAT TAC TTC CTA CAG GAG Val Lys Lys Gln Ser Gln Leu Ser Leu Leu Asp Tyr Phe Leu Gln Glu 625 630 635	2342
CAC GGC AGT TAC ACC ACT GAG GCA TTC CTC AGT GCA CAG CGC AAT TTT His Gly Ser Tyr Thr Thr Glu Ala Phe Leu Ser Ala Gln Arg Asn Phe 640 645 650	2390
GTG CAA AGT TGT GCT GGG TAC TGC TTG GTC TGC TAC CTG CTG CAA GTC Val Gln Ser Cys Ala Gly Tyr Cys Leu Val Cys Tyr Leu Leu Gln Val 655 660 665 670	2438
AAG GAC AGA CAC AAT GGG AAT ATC CTT TTG GAC GCA GAA GGC CAC ATC Lys Asp Arg His Asn Gly Asn Ile Leu Leu Asp Ala Glu Gly His Ile 675 680 685	2486
ATC CAC ATC GAC TTT GGC TTC ATC CTC TCC AGC TCA CCC CGA AAT CTG Ile His Ile Asp Phe Gly Phe Ile Leu Ser Ser Ser Pro Arg Asn Leu 690 695 700	2534
GGC TTT GAG ACG TCA GCC TTT AAG CTG ACC ACA GAG TTT GTG GAT GTG Gly Phe Glu Thr Ser Ala Phe Lys Leu Thr Thr Glu Phe Val Asp Val 705 710 715	2582
ATG GGC GGC CTG GAT GGC GAC ATG TTC AAC TAC TAT AAG ATG CTG ATG Met Gly Gly Leu Asp Gly Asp Met Phe Asn Tyr Tyr Lys Met Leu Met 720 725 730	2630
CTG CAA GGG CTG ATT GCC GCT CGG AAA CAC ATG GAC AAG GTG GTG CAG Leu Gln Gly Leu Ile Ala Ala Arg Lys His Met Asp Lys Val Val Gln 735 740 745 750	2678
ATC GTG GAG ATC ATG CAG CAA GGT TCT CAG CTT CCT TGC TTC CAT GGC Ile Val Glu Ile Met Gln Gln Gly Ser Gln Leu Pro Cys Phe His Gly 755 760 765	2726
TCC AGC ACC ATT CGA AAC CTC AAA GAG AGG TTC CAC ATG AGC ATG ACT Ser Ser Thr Ile Arg Asn Leu Lys Glu Arg Phe His Met Ser Met Thr 770 775 780	2774
GAG GAG CAG CTG CAG CTG CTG GTG GAG CAG ATG GTG GAT GGC AGT ATG Glu Glu Gln Leu Gln Leu Leu Val Glu Gln Met Val Asp Gly Ser Met 785 790 795	2822
CGG TCT ATC ACC ACC AAA CTC TAT GAC GGC TTC CAG TAC CTC ACC AAC	2870

Arg	Ser	Ile	Thr	Thr	Lys	Leu	Tyr	Asp	Gly	Phe	Gln	Tyr	Leu	Thr	Asn
800						805					810				

GGC ATC ATG TGA CACGCTCCTC AGCCCAGGAG TGGTGGGGGG TCCAGGGCAC	2922
Gly Ile Met *	
815	

CCTCCCTAGA GGGCCCTTGT CTGAGAAACC CCAAACCAGG AAACCCACAC TACCCAACCA	2982
TCCACCCAAG GGAAATGGAA GGCAAGAAAC ACGAAGGATC ATGTGGTAAC TGCGAGAGCT	3042
TGCTGAGGGG TGGGAGAGCC AGCTGTGGGG TCCAGACTTG TTGGGGCTTC CCTGCCCCCTC	3102
CTGGTCTGTG TCAGTATTAC CACCAGACTG ACTCCAGGAC TCACTGCCCT CCAGAAAACA	3162
GAGGTGACAA ATGTGAGGGA CACTGGGGCC TTCTTCTCC TTGTAGGGGT CTCTCAGAGG	3222
TTCTTTCCAC AGGCCATCCT CTTATTCCGT TCTGGGGCCC AGGAAGTGGG GAAGAGTAGG	3282
TTCTCGGTAC TTAGGACTTG ATCCTGTGGT TGCCACTGGC CATGCTGCTG CCCAGCTCTA	3342
CCCCTCCCAG GGACCTACCC CTCCCAGGGA CCGACCCCTG GCCCAAGCTC CCCTTGCTGG	3402
CGGGCGCTGC GTGGGCCCTG CACTTGCTGA GGTTCCCCAT CATGGGCAAG GCAAGGGAAT	3462
TCCCACAGCC CTCCAGTGTA CTGAGGGTAC TGGCCTAGCC ATGTGGAATT CCTTACCCTG	3522
ACTCCTTCCC CAAACCCAGG GAAAAGAGCT CTCAATTTTT TATTTTAAAT TTTTGTTGA	3582
AATAAAGTCC TTAGTTAGCC	3602

[0342]

SEQ ID NO:31

SEQUENCE CHARACTERISTICS:

LENGTH: 829 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met Arg Phe Leu Glu Ala Arg Ser Leu Ala Val Ala Met Gly Asp Thr

[illegible]

Pro Asp Thr Gly Leu Ser Pro Ser Lys Arg Thr His Gln Arg Ser Lys
275 280 285

Ser Asp Ala Thr Ala Ser Ile Ser Leu Ser Ser Asn Leu Lys Arg Thr
290 295 300

Ala Ser Asn Pro Lys Val Glu Asn Glu Asp Glu Glu Leu Ser Ser Ser
305 310 315 320

Thr Glu Ser Ile Asp Asn Ser Phe Ser Ser Pro Val Arg Leu Ala Pro
325 330 335

Glu Arg Glu Phe Ile Lys Ser Leu Met Ala Ile Gly Lys Arg Leu Ala
340 345 350

Thr Leu Pro Thr Lys Glu Gln Lys Thr Gln Arg Leu Ile Ser Glu Leu
355 360 365

Ser Leu Leu Asn His Lys Leu Pro Ala Arg Val Trp Leu Pro Thr Ala
370 375 380

Gly Phe Asp His His Val Val Arg Val Pro His Thr Gln Ala Val Val
385 390 395 400

Leu Asn Ser Lys Asp Lys Ala Pro Tyr Leu Ile Tyr Val Glu Val Leu
405 410 415

Glu Cys Glu Asn Phe Asp Thr Thr Ser Val Pro Ala Arg Ile Pro Glu
420 425 430

Asn Arg Ile Arg Ser Thr Arg Ser Val Glu Asn Leu Pro Glu Cys Gly
435 440 445

Ile Thr His Glu Gln Arg Ala Gly Ser Phe Ser Thr Val Pro Asn Tyr
450 455 460

Asp Asn Asp Asp Glu Ala Trp Ser Val Asp Asp Ile Gly Glu Leu Gln
465 470 475 480

Val Glu Leu Pro Glu Val His Thr Asn Ser Cys Asp Asn Ile Ser Gln
485 490 495

Phe Ser Val Asp Ser Ile Thr Ser Gln Glu Ser Lys Glu Pro Val Phe
500 505 510

Ile Ala Ala Gly Asp Ile Arg Arg Arg Leu Ser Glu Gln Leu Ala His
515 520 525

Thr Pro Thr Ala Phe Lys Arg Asp Pro Glu Asp Pro Ser Ala Val Ala

530		535		540
Leu Lys Glu Pro Trp Gln Glu Lys Val Arg Arg Ile Arg Glu Gly Ser				
545		550		555 560
Pro Tyr Gly His Leu Pro Asn Trp Arg Leu Leu Ser Val Ile Val Lys				
	565		570	575
Cys Gly Asp Asp Leu Arg Gln Glu Leu Leu Ala Phe Gln Val Leu Lys				
	580		585	590
Gln Leu Gln Ser Ile Trp Glu Gln Glu Arg Val Pro Leu Trp Ile Lys				
	595		600	605
Pro Ile Gln Asp Ser Cys Glu Ile Thr Thr Asp Ser Gly Met Ile Glu				
	610		615	620
Pro Val Val Asn Ala Val Ser Ile His Gln Val Lys Lys Gln Ser Gln				
	625		630	635 640
Leu Ser Leu Leu Asp Tyr Phe Leu Gln Glu His Gly Ser Tyr Thr Thr				
	645		650	655
Glu Ala Phe Leu Ser Ala Gln Arg Asn Phe Val Gln Ser Cys Ala Gly				
	660		665	670
Tyr Cys Leu Val Cys Tyr Leu Leu Gln Val Lys Asp Arg His Asn Gly				
	675		680	685
Asn Ile Leu Leu Asp Ala Glu Gly His Ile Ile His Ile Asp Phe Gly				
	690		695	700
Phe Ile Leu Ser Ser Ser Pro Arg Asn Leu Gly Phe Glu Thr Ser Ala				
	705		710	715 720
Phe Lys Leu Thr Thr Glu Phe Val Asp Val Met Gly Gly Leu Asp Gly				
	725		730	735
Asp Met Phe Asn Tyr Tyr Lys Met Leu Met Leu Gln Gly Leu Ile Ala				
	740		745	750
Ala Arg Lys His Met Asp Lys Val Val Gln Ile Val Glu Ile Met Gln				
	755		760	765
Gln Gly Ser Gln Leu Pro Cys Phe His Gly Ser Ser Thr Ile Arg Asn				
	770		775	780
Leu Lys Glu Arg Phe His Met Ser Met Thr Glu Glu Gln Leu Gln Leu				
	785		790	795 800

Leu Val Glu Gln Met Val Asp Gly Ser Met Arg Ser Ile Thr Thr Lys
805 810 815

Leu Tyr Asp Gly Phe Gln Tyr Leu Thr Asn Gly Ile Met
820 825

[0343]

SEQ ID NO:32

SEQUENCE CHARACTERISTICS:

LENGTH: 2487 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGAGATTCT TGGAAGCTCG AAGTCTGGCT GTGGCCATGG GAGATACAGT AGTGGAGCCT	60
GCCCCCTTGA AGCCAACTTC TGAGCCCACT TCTGGCCCAC CAGGGAATAA TGGGGGGTCC	120
CTGCTAAGTG TCATCACGGA GGGGGTCGGG GAACTATCAG TGATTGACCC TGAGGTGGCC	180
CAGAAGGCCT GCCAGGAGGT GTTGGAGAAA GTCAAGCTTT TGCATGGAGG CGTGGCAGTC	240
TCTAGCAGAG GCACCCCACT GGAGTTGGTC AATGGGGATG GTGTGGACAG TGAGATCCGT	300
TGCCTAGATG ATCCACCTGC CCAGATCAGG GAGGAGGAAG ATGAGATGGG GGCCGCTGTG	360
GCCTCAGGCA CAGCCAAAGG AGCAAGAAGA CGGCGGCAGA ACAACTCAGC TAAACAGTCT	420
TGGCTGCTGA GGCTGTTTGA GTCAAAACTG TTTGACATCT CCATGGCCAT TTCATACCTG	480
TATAACTCCA AGGAGCCTGG AGTACAAGCC TACATTGGCA ACCGGCTCTT CTGCTTTTCGC	540
AACGAGGACG TGGACTTCTA TCTGCCCCAG TTGCTTAACA TGTACATCCA CATGGATGAG	600
GACGTGGGTG ATGCCATTAA GCCCTACATA GTCCACCGTT GCCGCCAGAG CATTAACTTT	660
TCCCTCCAGT GTGCCCTGTT GCTTGGGGCC TATCTTTCAG ACATGCACAT TTCCACTCAA	720

CGACACTCCC GIGGGACCAA GCTACGGAAG CTGATCCTCT CAGATGAGCT AAAGCCAGCT	780
CACAGGAAGA GGGAGCTGCC CTCCTTGAGC CCGGCCCCCTG ATACAGGGCT GTCTCCCTCC	840
AAAAGGACTC ACCAGCGCTC TAAGTCAGAT GCCACTGCCA GCATAAGTCT CAGCAGCAAC	900
CTGAAACGAA CAGCCAGCAA CCCTAAAGTG GAGAATGAGG ATGAGGAGCT CTCCTCCAGC	960
ACCGAGAGTA TTGATAATTC ATTTCAGTTCC CCTGTTTCGAC TGGCTCCTGA GAGAGAATTC	1020
ATCAAGTCCC TGATGGCGAT CGGCAAGCGG CTGGCCACGC TCCCCACCAA AGAGCAGAAA	1080
ACACAGAGGC TGATCTCAGA GCTCTCCCTG CTCAACCATA AGCTCCCTGC CCGAGTCTGG	1140
CTGCCCCTG CTGGCTTTGA CCACCACGTG GTCCGTGTAC CCCACACACA GGCTGTTGTC	1200
CTCAACTCCA AGGACAAGGC TCCCTACCTG ATTTATGTGG AAGTCCTTGA ATGTGAAAAC	1260
TTTGACACCA CCAGTGTCCC TGCCCGGATC CCCGAGAACC GAATTCGGAG TACGAGGTCC	1320
GTAGAAACT TGCCCGAATG TGGTATTACC CATGAGCAGC GAGCTGGCAG CTTTCAGCACT	1380
GTGCCCCAAT ATGACAACGA TGATGAGGCC TGGTCGGTGG ATGACATAGG CGAGCTGCAA	1440
GTGGAGCTCC CCGAAGTGCA TACCAACAGC TGTGACAACA TCTCCCAGTT CTCTGTGGAC	1500
AGCATCACCA GCCAGGAGAG CAAGGAGCCT GTGTTTCAATG CAGCAGGGGA CATCCGCCGG	1560
CGCCTTTTCGG AACAGCTGGC TCATACCCCG ACAGCCTTCA AACGAGACCC AGAAGATCCT	1620
TCTGCAGTTG CTCTCAAAGA GCCCTGGCAG GAGAAAGTAC GGCGGATCAG AGAGGGCTCC	1680
CCCTACGGCC ATCTCCCCAA TTGGCGGCTC CTGTCAGTCA TTGTCAAGTG TGGGGATGAC	1740
CTTCGGCAAG AGCTTCTGGC CTTTCAGGTG TTGAAGCAAC TGCAGTCCAT TTGGGAACAG	1800
GAGCGAGTGC CCCTTTGGAT CAAGCCAATA CAAGATTCTT GTGAAATTAC GACTGATAGT	1860
GGCATGATTG AACCAGTGGT CAATGCTGTG TCCATCCATC AGGTGAAGAA ACAGTCACAG	1920
CTCTCCTTGC TCGATTACTT CCTACAGGAG CACGGCAGTT ACACCACTGA GGCATTCTC	1980
AGTGCACAGC GCAATTTTGT GCAAAGTGT GCTGGGTACT GCTTGGTCTG CTACCTGCTG	2040
CAAGTCAAGG ACAGACACAA TGGGAATATC CTTTTGGACG CAGAAGGCCA CATCATCCAC	2100
ATCGACTTTG GCTTCATCCT CTCCAGCTCA CCCCAGAAATC TGGGCTTTGA GACGTCAGCC	2160
TTTAAGCTGA CCACAGAGTT TGTGGATGTG ATGGGCGGCC TGGATGGCGA CATGTTCAAC	2220

-170-

TACTATAAGA TGCTGATGCT GCAAGGGCTG ATTGCCGCTC GGAAACACAT GGACAAGGTG 2280
GTGCAGATCG TGGAGATCAT GCAGCAAGGT TCTCAGCTTC CTTGCTTCCA TGGCTCCAGC 2340
ACCATTTCGAA ACCTCAAAGA GAGGTTCCAC ATGAGCATGA CTGAGGAGCA GCTGCAGCTG 2400
CTGGTGGAGC AGATGGTGGG TGGCAGTATG CCGTCTATCA CCACCAAACCT CTATGACGGC 2460
TTCCAGTACC TCACCAACGG CATCATG 2487

[0344]

SEQ ID NO:33

SEQUENCE CHARACTERISTICS:

LENGTH: 3324 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-428B12c1

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 115..2601

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

CCGGAATTCC GGAAGGCCG GAGCAAGTTT TGAAGAAGTC CCTATCAGAT TAACTTGGT 60
TGACTACTCC GGAGCAGCCA CTAAGAGGGA TGAACAGGCC TCGTGGAAA TTGA ATG 117
Met
1

AGA	TTC	TTG	GAA	GCT	CGA	AGT	CTG	GCT	GIG	GCC	ATG	GGA	GAT	ACA	GTA	165
Arg	Phe	Leu	Glu	Ala	Arg	Ser	Leu	Ala	Val	Ala	Met	Gly	Asp	Thr	Val	
			5					10					15			
GTG	GAG	CCT	GCC	CCC	TTG	AAG	CCA	ACT	TCT	GAG	CCC	ACT	TCT	GGC	CCA	213
Val	Glu	Pro	Ala	Pro	Leu	Lys	Pro	Thr	Ser	Glu	Pro	Thr	Ser	Gly	Pro	
		20					25					30				
CCA	GGG	AAT	AAT	GGG	GGG	TCC	CTG	CTA	AGT	GTC	ATC	ACG	GAG	GGG	GTC	261
Pro	Gly	Asn	Asn	Gly	Gly	Ser	Leu	Leu	Ser	Val	Ile	Thr	Glu	Gly	Val	
	35					40					45					
GGG	GAA	CTA	TCA	GTG	ATT	GAC	CCT	GAG	GTG	GCC	CAG	AAG	GCC	TGC	CAG	309
Gly	Glu	Leu	Ser	Val	Ile	Asp	Pro	Glu	Val	Ala	Gln	Lys	Ala	Cys	Gln	
50					55					60					65	
GAG	GTG	TTG	GAG	AAA	GTC	AAG	CTT	TTG	CAT	GGA	GGC	GTG	GCA	GTC	TCT	357
Glu	Val	Leu	Glu	Lys	Val	Lys	Leu	Leu	His	Gly	Gly	Val	Ala	Val	Ser	
				70					75					80		
AGC	AGA	GGC	ACC	CCA	CTG	GAG	TTG	GTC	AAT	GGG	GAT	GGT	GTG	GAC	AGT	405
Ser	Arg	Gly	Thr	Pro	Leu	Glu	Leu	Val	Asn	Gly	Asp	Gly	Val	Asp	Ser	
			85					90					95			
GAG	ATC	CGT	TGC	CTA	GAT	GAT	CCA	CCT	GCC	CAG	ATC	AGG	GAG	GAG	GAA	453
Glu	Ile	Arg	Cys	Leu	Asp	Asp	Pro	Pro	Ala	Gln	Ile	Arg	Glu	Glu	Glu	
		100					105					110				
GAT	GAG	ATG	GGG	GCC	GCT	GTG	GCC	TCA	GGC	ACA	GCC	AAA	GGA	GCA	AGA	501
Asp	Glu	Met	Gly	Ala	Ala	Val	Ala	Ser	Gly	Thr	Ala	Lys	Gly	Ala	Arg	
	115					120					125					
AGA	CGG	CGG	CAG	AAC	AAC	TCA	GCT	AAA	CAG	TCT	TGG	CTG	CTG	AGG	CTG	549
Arg	Arg	Arg	Gln	Asn	Asn	Ser	Ala	Lys	Gln	Ser	Trp	Leu	Leu	Arg	Leu	
130					135					140					145	
TTT	GAG	TCA	AAA	CTG	TTT	GAC	ATC	TCC	ATG	GCC	ATT	TCA	TAC	CTG	TAT	597
Phe	Glu	Ser	Lys	Leu	Phe	Asp	Ile	Ser	Met	Ala	Ile	Ser	Tyr	Leu	Tyr	
			150					155						160		
AAC	TCC	AAG	GAG	CCT	GGA	GTA	CAA	GCC	TAC	ATT	GGC	AAC	CGG	CTC	TTC	645
Asn	Ser	Lys	Glu	Pro	Gly	Val	Gln	Ala	Tyr	Ile	Gly	Asn	Arg	Leu	Phe	
			165					170					175			
TGC	TTT	CGC	AAC	GAG	GAC	GTG	GAC	TTC	TAT	CTG	CCC	CAG	TTG	CTT	AAC	693
Cys	Phe	Arg	Asn	Glu	Asp	Val	Asp	Phe	Tyr	Leu	Pro	Gln	Leu	Leu	Asn	
		180					185					190				
ATG	TAC	ATC	CAC	ATG	GAT	GAG	GAC	GTG	GGT	GAT	GCC	ATT	AAG	CCC	TAC	741

Met	Tyr	Ile	His	Met	Asp	Glu	Asp	Val	Gly	Asp	Ala	Ile	Lys	Pro	Tyr	
195						200					205					
ATA	GTC	CAC	CGT	TGC	CGC	CAG	AGC	ATT	AAC	TTT	TCC	CTC	CAG	TGT	GCC	789
Ile	Val	His	Arg	Cys	Arg	Gln	Ser	Ile	Asn	Phe	Ser	Leu	Gln	Cys	Ala	
210					215				220						225	
CTG	TTG	CTT	GGG	GCC	TAT	TCT	TCA	GAC	ATG	CAC	ATT	TCC	ACT	CAA	CGA	837
Leu	Leu	Leu	Gly	Ala	Tyr	Ser	Ser	Asp	Met	His	Ile	Ser	Thr	Gln	Arg	
				230					235					240		
CAC	TCC	CGT	GGG	ACC	AAG	CTA	CGG	AAG	CTG	ATC	CTC	TCA	GAT	GAG	CTA	885
His	Ser	Arg	Gly	Thr	Lys	Leu	Arg	Lys	Leu	Ile	Leu	Ser	Asp	Glu	Leu	
			245					250					255			
AAG	CCA	GCT	CAC	AGG	AAG	AGG	GAG	CTG	CCC	TCC	TTG	AGC	CCG	GCC	CCT	933
Lys	Pro	Ala	His	Arg	Lys	Arg	Glu	Leu	Pro	Ser	Leu	Ser	Pro	Ala	Pro	
		260					265					270				
GAT	ACA	GGG	CTG	TCT	CCC	TCC	AAA	AGG	ACT	CAC	CAG	CGC	TCT	AAG	TCA	981
Asp	Thr	Gly	Leu	Ser	Pro	Ser	Lys	Arg	Thr	His	Gln	Arg	Ser	Lys	Ser	
	275					280					285					
GAT	GCC	ACT	GCC	AGC	ATA	AGT	CTC	AGC	AGC	AAC	CTG	AAA	CGA	ACA	GCC	1029
Asp	Ala	Thr	Ala	Ser	Ile	Ser	Leu	Ser	Ser	Asn	Leu	Lys	Arg	Thr	Ala	
290					295					300					305	
AGC	AAC	CCT	AAA	GTG	GAG	AAT	GAG	GAT	GAG	GAG	CTC	TCC	TCC	AGC	ACC	1077
Ser	Asn	Pro	Lys	Val	Glu	Asn	Glu	Asp	Glu	Glu	Leu	Ser	Ser	Ser	Thr	
				310					315					320		
GAG	AGT	ATT	GAT	AAT	TCA	TTC	AGT	TCC	CCT	GTT	CGA	CTG	GCT	CCT	GAG	1125
Glu	Ser	Ile	Asp	Asn	Ser	Phe	Ser	Ser	Pro	Val	Arg	Leu	Ala	Pro	Glu	
			325					330					335			
AGA	GAA	TTC	ATC	AAG	TCC	CTG	ATG	GCG	ATC	GGC	AAG	CGG	CTG	GCC	ACG	1173
Arg	Glu	Phe	Ile	Lys	Ser	Leu	Met	Ala	Ile	Gly	Lys	Arg	Leu	Ala	Thr	
		340					345					350				
CTC	CCC	ACC	AAA	GAG	CAG	AAA	ACA	CAG	AGG	CTG	ATC	TCA	GAG	CTC	TCC	1221
Leu	Pro	Thr	Lys	Glu	Gln	Lys	Thr	Gln	Arg	Leu	Ile	Ser	Glu	Leu	Ser	
		355				360					365					
CTG	CTC	AAC	CAT	AAG	CTC	CCT	GCC	CGA	GTC	TGG	CTG	CCC	ACT	GCT	GGC	1269
Leu	Leu	Asn	His	Lys	Leu	Pro	Ala	Arg	Val	Trp	Leu	Pro	Thr	Ala	Gly	
370					375					380					385	
TTT	GAC	CAC	CAC	GTG	GTC	CGT	GTA	CCC	CAC	ACA	CAG	GCT	GTT	GTC	CTC	1317
Phe	Asp	His	His	Val	Val	Arg	Val	Pro	His	Thr	Gln	Ala	Val	Val	Leu	

390	395	400	
AAC TCC AAG GAC AAG GCT CCC TAC CTG ATT TAT GTG GAA GTC CTT GAA Asn Ser Lys Asp Lys Ala Pro Tyr Leu Ile Tyr Val Glu Val Leu Glu 405 410 415			1365
TGT GAA AAC TTT GAC ACC ACC AGT GTC CCT GCC CGG ATC CCC GAG AAC Cys Glu Asn Phe Asp Thr Thr Ser Val Pro Ala Arg Ile Pro Glu Asn 420 425 430			1413
CGA ATT CGG AGT ACG AGG TCC GTA GAA AAC TTG CCC GAA TGT GGT ATT Arg Ile Arg Ser Thr Arg Ser Val Glu Asn Leu Pro Glu Cys Gly Ile 435 440 445			1461
ACC CAT GAG CAG CGA GCT GGC AGC TTC AGC ACT GTG CCC AAC TAT GAC Thr His Glu Gln Arg Ala Gly Ser Phe Ser Thr Val Pro Asn Tyr Asp 450 455 460 465			1509
AAC GAT GAT GAG GCC TGG TCG GTG GAT GAC ATA GGC GAG CTG CAA GTG Asn Asp Asp Glu Ala Trp Ser Val Asp Asp Ile Gly Glu Leu Gln Val 470 475 480			1557
GAG CTC CCC GAA GTG CAT ACC AAC AGC TGT GAC AAC ATC TCC CAG TTC Glu Leu Pro Glu Val His Thr Asn Ser Cys Asp Asn Ile Ser Gln Phe 485 490 495			1605
TCT GTG GAC AGC ATC ACC AGC CAG GAG AGC AAG GAG CCT GTG TTC ATT Ser Val Asp Ser Ile Thr Ser Gln Glu Ser Lys Glu Pro Val Phe Ile 500 505 510			1653
GCA GCA GGG GAC ATC CGC CGG CGC CTT TCG GAA CAG CTG GCT CAT ACC Ala Ala Gly Asp Ile Arg Arg Arg Leu Ser Glu Gln Leu Ala His Thr 515 520 525			1701
CCG ACA GCC TTC AAA CGA GAC CCA GAA GAT CCT TCT GCA GTT GCT CTC Pro Thr Ala Phe Lys Arg Asp Pro Glu Asp Pro Ser Ala Val Ala Leu 530 535 540 545			1749
AAA GAG CCC TGG CAG GAG AAA GTA CGG CGG ATC AGA GAG GGC TCC CCC Lys Glu Pro Trp Gln Glu Lys Val Arg Arg Ile Arg Glu Gly Ser Pro 550 555 560			1797
TAC GGC CAT CTC CCC AAT TGG CGG CTC CTG TCA GTC ATT GTC AAG TGT Tyr Gly His Leu Pro Asn Trp Arg Leu Leu Ser Val Ile Val Lys Cys 565 570 575			1845
GGG GAT GAC CTT CGG CAA GAG CTT CTG GCC TTT CAG GTG TTG AAG CAA Gly Asp Asp Leu Arg Gln Glu Leu Leu Ala Phe Gln Val Leu Lys Gln 580 585 590			1893

CTG CAG TCC ATT TGG GAA CAG GAG CGA GTG CCC CTT TGG ATC AAG CCA Leu Gln Ser Ile Trp Glu Gln Glu Arg Val Pro Leu Trp Ile Lys Pro 595 600 605	1941
ATA CAA GAT TCT TGT GAA ATT ACG ACT GAT AGT GGC ATG ATT GAA CCA Ile Gln Asp Ser Cys Glu Ile Thr Thr Asp Ser Gly Met Ile Glu Pro 610 615 620 625	1989
GTG GTC AAT GCT GTG TCC ATC CAT CAG GTG AAG AAA CAG TCA CAG CTC Val Val Asn Ala Val Ser Ile His Gln Val Lys Lys Gln Ser Gln Leu 630 635 640	2037
TCC TTG CTC GAT TAC TTC CTA CAG GAG CAC GGC AGT TAC ACC ACT GAG Ser Leu Leu Asp Tyr Phe Leu Gln Glu His Gly Ser Tyr Thr Thr Glu 645 650 655	2085
GCA TTC CTC AGT GCA CAG CGC AAT TTT GTG CAA AGT TGT GCT GGG TAC Ala Phe Leu Ser Ala Gln Arg Asn Phe Val Gln Ser Cys Ala Gly Tyr 660 665 670	2133
TGC TTG GTC TGC TAC CTG CTG CAA GTC AAG GAC AGA CAC AAT GGG AAT Cys Leu Val Cys Tyr Leu Leu Gln Val Lys Asp Arg His Asn Gly Asn 675 680 685	2181
ATC CTT TTG GAC GCA GAA GGC CAC ATC ATC CAC ATC GAC TTT GGC TTC Ile Leu Leu Asp Ala Glu Gly His Ile Ile His Ile Asp Phe Gly Phe 690 695 700 705	2229
ATC CTC TCC AGC TCA CCC CGA AAT CTG GGC TTT GAG ACG TCA GCC TTT Ile Leu Ser Ser Ser Pro Arg Asn Leu Gly Phe Glu Thr Ser Ala Phe 710 715 720	2277
AAG CTG ACC ACA GAG TTT GTG GAT GTG ATG GGC GGC CTG GAT GGC GAC Lys Leu Thr Thr Glu Phe Val Asp Val Met Gly Gly Leu Asp Gly Asp 725 730 735	2325
ATG TTC AAC TAC TAT AAG ATG CTG ATG CTG CAA GGC CTG ATT GCC GCT Met Phe Asn Tyr Tyr Lys Met Leu Met Leu Gln Gly Leu Ile Ala Ala 740 745 750	2373
CGG AAA CAC ATG GAC AAG GTG GTG CAG ATC GTG GAG ATC ATG CAG CAA Arg Lys His Met Asp Lys Val Val Gln Ile Val Glu Ile Met Gln Gln 755 760 765	2421
GGT TCT CAG CTT CCT TGC TTC CAT GGC TCC AGC ACC ATT CGA AAC CTC Gly Ser Gln Leu Pro Cys Phe His Gly Ser Ser Thr Ile Arg Asn Leu 770 775 780 785	2469
AAA GAG AGG TTC CAC ATG AGC ATG ACT GAG GAG CAG CTG CAG CTG CTG	2517

Lys	Glu	Arg	Phe	His	Met	Ser	Met	Thr	Glu	Glu	Gln	Leu	Gln	Leu	Leu	
				790					795					800		
GTG	GAG	CAG	ATG	GTG	GAT	GGC	AGT	ATG	CGG	TCT	ATC	ACC	ACC	AAA	CTC	2565
Val	Glu	Gln	Met	Val	Asp	Gly	Ser	Met	Arg	Ser	Ile	Thr	Thr	Lys	Leu	
			805					810						815		
TAT	GAC	GGC	TTC	CAG	TAC	CTC	ACC	AAC	GGC	ATC	ATG	TGA	CACGCTCCTC			2614
Tyr	Asp	Gly	Phe	Gln	Tyr	Leu	Thr	Asn	Gly	Ile	Met	*				
		820					825					830				
AGCCCAGGAG	TGGTGGGGGG	TCCAGGGCAC	CCTCCCTAGA	GGGCCCTTGT	CTGAGAAACC											2674
CCAAACCAGG	AAACCCACCC	TACCCAACCA	TCCACCCAAG	GGAAATGGAA	GGCAAGAAAC											2734
ACGAAGGATC	ATGTGGTAAC	TGCGAGAGCT	TGCTGAGGGG	TGGGAGAGCC	AGCTGTGGGG											2794
TCCAGACTTG	TTGGGGCTTC	CCTGCCCCCTC	CTGGTCTGIG	TCAGTATTAC	CACCAGACTG											2854
ACTCCAGGAC	TCACTGCCCT	CCAGAAAACA	GAGGTGACAA	ATGTGAGGGA	CAC'TGGGGCC											2914
TTTCTTCTCC	TTGTAGGGGT	CTCTCAGAGG	TTCTTTCCAC	AGGCCATCCT	CTTATTCGGT											2974
TCTGGGGCCC	AGGAAGTGGG	GAAGAGTAGG	TTCTCGGTAC	TTAGGACTTG	ATCCTGTGGT											3034
TGCCACTGGC	CATGCTGCTG	CCCAGCTCTA	CCCCTCCCAG	GGACCTACCC	CTCCCAGGGA											3094
CCGACCCCTG	GCCCAAGCTC	CCCTTGCTGG	CGGGCGCTGC	GTGGGCCCTG	CAC'TTGCTGA											3154
GGTTCCCCAT	CATGGGCAAG	GCAAGGGAAT	TCCCACAGCC	CTCCAGTGTA	CTGAGGGTAC											3214
TGGCCTAGCC	ATGTGGAATT	CCCTACCCTG	ACTCCTTCCC	CAAACCCAGG	GAAAAGAGCT											3274
CTCAATTTTT	TATTTTAAAT	TTTGTGTTGA	AATAAAGTCC	TTAGTTAGCC												3324

[0345]

SEQ ID NO:34

SEQUENCE CHARACTERISTICS:

LENGTH: 810 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met	Pro	Met	Asp	Leu	Ile	Leu	Val	Val	Trp	Phe	Cys	Val	Cys	Thr	Ala	1	5	10	15
Arg	Thr	Val	Val	Gly	Phe	Gly	Met	Asp	Pro	Asp	Leu	Gln	Met	Asp	Ile	20	25	30	
Val	Thr	Glu	Leu	Asp	Leu	Val	Asn	Thr	Thr	Leu	Gly	Val	Ala	Gln	Val	35	40	45	
Ser	Gly	Met	His	Asn	Ala	Ser	Lys	Ala	Phe	Leu	Phe	Gln	Asp	Ile	Glu	50	55	60	
Arg	Glu	Ile	His	Ala	Ala	Pro	His	Val	Ser	Glu	Lys	Leu	Ile	Gln	Leu	65	70	75	80
Phe	Gln	Asn	Lys	Ser	Glu	Phe	Thr	Ile	Leu	Ala	Thr	Val	Gln	Gln	Lys	85	90	95	
Pro	Ser	Thr	Ser	Gly	Val	Ile	Leu	Ser	Ile	Arg	Glu	Leu	Glu	His	Ser	100	105	110	
Tyr	Phe	Glu	Leu	Glu	Ser	Ser	Gly	Leu	Arg	Asp	Glu	Ile	Arg	Tyr	His	115	120	125	
Tyr	Ile	His	Asn	Gly	Lys	Pro	Arg	Thr	Glu	Ala	Leu	Pro	Tyr	Arg	Met	130	135	140	
Ala	Asp	Gly	Gln	Trp	His	Lys	Val	Ala	Leu	Ser	Val	Ser	Ala	Ser	His	145	150	155	160
Leu	Leu	Leu	His	Val	Asp	Cys	Asn	Arg	Ile	Tyr	Glu	Arg	Val	Ile	Asp	165	170	175	
Pro	Pro	Asp	Thr	Asn	Leu	Pro	Pro	Gly	Ile	Asn	Leu	Trp	Leu	Gly	Gln	180	185	190	
Arg	Asn	Gln	Lys	His	Gly	Leu	Phe	Lys	Gly	Ile	Ile	Gln	Asp	Gly	Lys	195	200	205	
Ile	Ile	Phe	Met	Pro	Asn	Gly	Tyr	Ile	Thr	Gln	Cys	Pro	Asn	Leu	Asn	210	215	220	
His	Thr	Cys	Pro	Thr	Cys	Ser	Asp	Phe	Leu	Ser	Leu	Val	Gln	Gly	Ile	225	230	235	240
Met	Asp	Leu	Gln	Glu	Leu	Leu	Ala	Lys	Met	Thr	Ala	Lys	Leu	Asn	Tyr	245	250	255	

Ala Glu Thr Arg Leu Ser Gln Leu Glu Asn Cys His Cys Glu Lys Thr
260 265 270

Cys Gln Val Ser Gly Leu Leu Tyr Arg Asp Gln Asp Ser Trp Val Asp
275 280 285

Gly Asp His Cys Arg Asn Cys Thr Cys Lys Ser Gly Ala Val Glu Cys
290 295 300

Arg Arg Met Ser Cys Pro Pro Leu Asn Cys Ser Pro Asp Ser Leu Pro
305 310 315 320

Val His Ile Ala Gly Gln Cys Cys Lys Val Cys Arg Pro Lys Cys Ile
325 330 335

Tyr Gly Gly Lys Val Leu Ala Glu Gly Gln Arg Ile Leu Thr Lys Ser
340 345 350

Cys Arg Glu Cys Arg Gly Gly Val Leu Val Lys Ile Thr Glu Met Cys
355 360 365

Pro Pro Leu Asn Cys Ser Glu Lys Asp His Ile Leu Pro Glu Asn Gln
370 375 380

Cys Cys Arg Val Cys Arg Gly His Asn Phe Cys Ala Glu Gly Pro Lys
385 390 395 400

Cys Gly Glu Asn Ser Glu Cys Lys Asn Trp Asn Thr Lys Ala Thr Cys
405 410 415

Glu Cys Lys Ser Gly Tyr Ile Ser Val Gln Gly Asp Ser Ala Tyr Cys
420 425 430

Glu Asp Ile Asp Glu Cys Ala Ala Lys Met His Tyr Cys His Ala Asn
435 440 445

Thr Val Cys Val Asn Leu Pro Gly Leu Tyr Arg Cys Asp Cys Val Pro
450 455 460

Gly Tyr Ile Arg Val Asp Asp Phe Ser Cys Thr Glu His Asp Glu Cys
465 470 475 480

Gly Ser Gly Gln His Asn Cys Asp Glu Asn Ala Ile Cys Thr Asn Thr
485 490 495

Val Gln Gly His Ser Cys Thr Cys Lys Pro Gly Tyr Val Gly Asn Gly
500 505 510

Thr Ile Cys Arg Ala Phe Cys Glu Glu Gly Cys Arg Tyr Gly Gly Thr

515		520		525
Cys Val Ala Pro Asn Lys	Cys Val Cys Pro Ser Gly Phe Thr Gly Ser			
530	535	540		
His Cys Glu Lys Asp Ile Asp Glu Cys Ser Glu Gly Ile Ile Glu Cys				
545	550	555		560
His Asn His Ser Arg Cys Val Asn Leu Pro Gly Trp Tyr His Cys Glu				
	565	570		575
Cys Arg Ser Gly Phe His Asp Asp Gly Thr Tyr Ser Leu Ser Gly Glu				
	580	585		590
Ser Cys Ile Asp Ile Asp Glu Cys Ala Leu Arg Thr His Thr Cys Trp				
595	600	605		
Asn Asp Ser Ala Cys Ile Asn Leu Ala Gly Gly Phe Asp Cys Leu Cys				
610	615	620		
Pro Ser Gly Pro Ser Cys Ser Gly Asp Cys Pro His Glu Gly Gly Leu				
625	630	635		640
Lys His Asn Gly Gln Val Trp Thr Leu Lys Glu Asp Arg Cys Ser Val				
	645	650		655
Cys Ser Cys Lys Asp Gly Lys Ile Phe Cys Arg Arg Thr Ala Cys Asp				
	660	665		670
Cys Gln Asn Pro Ser Ala Asp Leu Phe Cys Cys Pro Glu Cys Asp Thr				
	675	680		685
Arg Val Thr Ser Gln Cys Leu Asp Gln Asn Gly His Lys Leu Tyr Arg				
690	695	700		
Ser Gly Asp Asn Trp Thr His Ser Cys Gln Gln Cys Arg Cys Leu Glu				
705	710	715		720
Gly Glu Val Asp Cys Trp Pro Leu Thr Cys Pro Asn Leu Ser Cys Glu				
	725	730		735
Tyr Thr Ala Ile Leu Glu Gly Glu Cys Cys Pro Arg Cys Val Ser Asp				
	740	745		750
Pro Cys Leu Ala Asp Asn Ile Thr Tyr Asp Ile Arg Lys Thr Cys Leu				
755	760	765		
Asp Ser Tyr Gly Val Ser Arg Leu Ser Gly Ser Val Trp Thr Met Ala				
770	775	780		

Gly Ser Pro Cys Thr Thr Cys Lys Cys Lys Asn Gly Arg Val Cys Cys
785 790 795 800

Ser Val Asp Phe Glu Cys Leu Gln Asn Asn
805 810

[0346]

SEQ ID NO:35

SEQUENCE CHARACTERISTICS:

LENGTH: 2430 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGCCGATGG ATTTGATTTT AGTGTGTGG TTCTGTGTGT GCACTGCCAG GACAGTGGTG	60
GGCTTTGGGA TGGACCCTGA CCTTCAGATG GATATCGTCA CCGAGCTTGA CCTTGTGAAC	120
ACCACCTTIG GAGTGTCTCA GGTGTCTGGA ATGCACAATG CCAGCAAAGC ATTTTATTTT	180
CAAGACATAG AAAGAGAGAT CCATGCAGCT CCTCATGTGA GTGAGAAATT AATTCAGCTG	240
TTCCAGAACA AGAGTGAATT CACCATTTTG GCCACTGTAC AGCAGAAGCC ATCCACTTCA	300
GGAGTGATAC TGTCATTTCG AGAACTGGAG CACAGCTATT TTGAACTGGA GAGCAGTGGC	360
CTGAGGGATG AGATTCGGTA TCACTACATA CACAATGGGA AGCCAAGGAC AGAGGCACTT	420
CCTTACCGCA TGGCAGATGG ACAATGGCAC AAGGTTGCAC TGTCAGTTAG CGCCTCTCAT	480
CTCCTGCTCC ATGTCGACTG TAACAGGATT TATGAGCGTG TGATAGACCC TCCAGATACC	540
AACCTTCCCC CAGGAATCAA TTTATGGCTT GGCCAGCGCA ACCAAAAGCA TGGCTTATTC	600
AAAGGGATCA TCCAAGATGG GAAGATCATC TTTATGCCGA ATGGATATAT AACACAGTGT	660
CCAAATCTAA ATCACACTTG CCCAACCTGC AGTGATTTCT TAAGCCTGGT GCAAGGAATA	720

ATGGATTAC	AAGAGCTTTT	GGCCAAGATG	ACTGCAAAAC	TAAATTATGC	AGAGACAAGA	780
CTTAGTCAAT	TGGAAACTG	TCATTGTGAG	AAGACTTGTC	AAGTGAGTGG	ACTGCTCTAT	840
CGAGATCAAG	ACTCTTGGGT	AGATGGTGAC	CATTGCAGGA	ACTGCACTTG	CAAAAGTGGT	900
GCCGTGGAAT	GCCGAAGGAT	GTCCTGTCCC	CCTCTCAATT	GCTCCCCAGA	CTCCCTCCCA	960
GTACACATTG	CTGGCCAGTG	CTGTAAGGTC	TGCCGACCAA	AATGTATCTA	TGGAGGAAAA	1020
GTTCTTGACAG	AAGGCCAGCG	GATTTTAACC	AAGAGCTGTC	GGGAATGCCG	AGGTGGAGTT	1080
TTAGTAAAAA	TTACAGAAAT	GTGTCCTCCT	TTGAACTGCT	CAGAAAAGGA	TCACATTCTT	1140
CCTGAGAATC	AGTGCTGCCG	TGTCGTAGA	GGTCATAACT	TTTGTGCAGA	AGGACCTAAA	1200
TGTGGTGAAA	ACTCAGAGTG	CAAAAACCTGG	AATACAAAAG	CTACTTGTTGA	GTGCAAGAGT	1260
GGTTACATCT	CTGTCCAGGG	AGACTCTGCC	TACTGTGAAG	ATATTGATGA	GTGTGCAGCT	1320
AAGATGCATT	ACTGTCATGC	CAATACTGTG	TGTGTCAACC	TTCCTGGGTT	ATATCGCTGT	1380
GACTGTGTCC	CAGGATACAT	TCGTGTGGAT	GACTTCTCTT	GTACAGAACA	CGATGAATGT	1440
GGCAGCGGCC	AGCACAACTG	TGATGAGAAT	GCCATCTGCA	CCAACACTGT	CCAGGGACAC	1500
AGCTGCACCT	GCAAACCGGG	CTACGTGGGG	AACGGGACCA	TCTGCAGAGC	TTTCTGTGAA	1560
GAGGGCTGCA	GATACGGTGG	AACGTGTGTG	GCTCCCAACA	AATGTGTCTG	TCCATCTGGA	1620
TTACACAGGAA	GCCACTGCGA	GAAAGATATT	GATGAATGTT	CAGAGGGAAT	CATTGAGTGC	1680
CACAACCAAT	CCCGCTGCGT	TAACCTGCCA	GGGTGGTACC	ACTGTGAGTG	CAGAAGCGGT	1740
TTCCATGACG	ATGGGACCTA	TTCACTGTCC	GGGGAGTCCT	GTATTGACAT	TGATGAATGT	1800
GCCTTAAGAA	CTCACACCTG	TTGGAACGAT	TCTGCCTGCA	TCAACCTGGC	AGGGGGTTTT	1860
GACTGTCTCT	GCCCCTCTGG	GCCCTCCTGC	TCTGGTGACT	GTCTCATGA	AGGGGGGCTG	1920
AAGCACAATG	GCCAGGTGTG	GACCTTGAAA	GAAGACAGGT	GTCTGTCTTG	CTCCTGCAAG	1980
GATGGCAAGA	TATCTTGCCG	ACGGACAGCT	TGTGATTGCC	AGAATCCAAG	TGCTGACCTA	2040
TTCTGTTGCC	CAGAAATGTA	CACCAGAGTC	ACAAGTCAAT	GTTTAGACCA	AAATGGTCAC	2100
AAGCTGTATC	GAAGTGGAGA	CAATTGGACC	CATAGCTGTC	AGCAGTGTG	GTGTCTGGAA	2160
GGAGAGGTAG	ATTGCTGGCC	ACTCACTTGC	CCCAACTTGA	GCTGTGAGTA	TACAGCTATC	2220

TTAGAAGGGG AATGTTGTC CCGCTGTGTC AGTGACCCCT GCCTAGCTGA TAACATCACC	2280
TATGACATCA GAAAACTTG CCTGGACAGC TATGGTGTTT CACGGCTTAG TGGCTCAGTG	2340
TGGACGATGG CTGGATCTCC CTGCACAACC TGTAATGCA AGAATGGAAG AGTCTGTTGT	2400
TCTGTGGATT TTGAGTGTCT TCAAATAAT	2430

[0347]

SEQ ID NO:36

SEQUENCE CHARACTERISTICS:

LENGTH: 2977 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-073E07

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 103..2532

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

TAGCAAGTTT GCGGCTCCA AGCCAGGCGC GCCTCAGGAT CCAGGCTCAT TTGCTTCCAC	60
CTAGCTTCGG TGCCCCCTGC TAGGCGGGGA CCCTCGAGAG CG ATG CCG ATG GAT	114
	Met Pro Met Asp
	1
TTG ATT TTA GTT GTG TGG TTC TGT GTG TGC ACT GCC AGG ACA GTG GTG	162
Leu Ile Leu Val Val Trp Phe Cys Val Cys Thr Ala Arg Thr Val Val	

5	10	15	20	
GGC TTT GGG ATG GAC CCT GAC CTT CAG ATG GAT ATC GTC ACC GAG CTT				210
Gly Phe Gly Met Asp Pro Asp Leu Gln Met Asp Ile Val Thr Glu Leu	25	30	35	
GAC CTT GTG AAC ACC ACC CTT GGA GTT GCT CAG GTG TCT GGA ATG CAC				258
Asp Leu Val Asn Thr Thr Leu Gly Val Ala Gln Val Ser Gly Met His	40	45	50	
AAT GCC AGC AAA GCA TTT TTA TTT CAA GAC ATA GAA AGA GAG ATC CAT				306
Asn Ala Ser Lys Ala Phe Leu Phe Gln Asp Ile Glu Arg Glu Ile His	55	60	65	
GCA GCT CCT CAT GTG AGT GAG AAA TTA ATT CAG CTG TTC CAG AAC AAG				354
Ala Ala Pro His Val Ser Glu Lys Leu Ile Gln Leu Phe Gln Asn Lys	70	75	80	
AGT GAA TTC ACC ATT TTG GCC ACT GTA CAG CAG AAG CCA TCC ACT TCA				402
Ser Glu Phe Thr Ile Leu Ala Thr Val Gln Gln Lys Pro Ser Thr Ser	85	90	95	100
GGA GTG ATA CTG TCC ATT CGA GAA CTG GAG CAC AGC TAT TTT GAA CTG				450
Gly Val Ile Leu Ser Ile Arg Glu Leu Glu His Ser Tyr Phe Glu Leu	105	110	115	
GAG AGC AGT GGC CTG AGG GAT GAG ATT CGG TAT CAC TAC ATA CAC AAT				498
Glu Ser Ser Gly Leu Arg Asp Glu Ile Arg Tyr His Tyr Ile His Asn	120	125	130	
GGG AAG CCA AGG ACA GAG GCA CTT CCT TAC CGC ATG GCA GAT GGA CAA				546
Gly Lys Pro Arg Thr Glu Ala Leu Pro Tyr Arg Met Ala Asp Gly Gln	135	140	145	
TGG CAC AAG GTT GCA CTG TCA GTT AGC GCC TCT CAT CTC CTG CTC CAT				594
Trp His Lys Val Ala Leu Ser Val Ser Ala Ser His Leu Leu Leu His	150	155	160	
GTC GAC TGT AAC AGG ATT TAT GAG CGT GTG ATA GAC CCT CCA GAT ACC				642
Val Asp Cys Asn Arg Ile Tyr Glu Arg Val Ile Asp Pro Pro Asp Thr	165	170	175	180
AAC CTT CCC CCA GGA ATC AAT TTA TGG CTT GGC CAG CGC AAC CAA AAG				690
Asn Leu Pro Pro Gly Ile Asn Leu Trp Leu Gly Gln Arg Asn Gln Lys	185	190	195	
CAT GGC TTA TTC AAA GGG ATC ATC CAA GAT GGG AAG ATC ATC TTT ATG				738
His Gly Leu Phe Lys Gly Ile Ile Gln Asp Gly Lys Ile Ile Phe Met	200	205	210	

CCG AAT GGA TAT ATA ACA CAG TGT CCA AAT CTA AAT CAC ACT TGC CCA	786
Pro Asn Gly Tyr Ile Thr Gln Cys Pro Asn Leu Asn His Thr Cys Pro	
215 220 225	
ACC TGC AGT GAT TTC TTA AGC CTG GTG CAA GGA ATA ATG GAT TTA CAA	834
Thr Cys Ser Asp Phe Leu Ser Leu Val Gln Gly Ile Met Asp Leu Gln	
230 235 240	
GAG CTT TTG GCC AAG ATG ACT GCA AAA CTA AAT TAT GCA GAG ACA AGA	882
Glu Leu Leu Ala Lys Met Thr Ala Lys Leu Asn Tyr Ala Glu Thr Arg	
245 250 255 260	
CTT AGT CAA TTG GAA AAC TGT CAT TGT GAG AAG ACT TGT CAA GTG AGT	930
Leu Ser Gln Leu Glu Asn Cys His Cys Glu Lys Thr Cys Gln Val Ser	
265 270 275	
GGA CTG CTC TAT CGA GAT CAA GAC TCT TGG GTA GAT GGT GAC CAT TGC	978
Gly Leu Leu Tyr Arg Asp Gln Asp Ser Trp Val Asp Gly Asp His Cys	
280 285 290	
AGG AAC TGC ACT TGC AAA AGT GGT GCC GTG GAA TGC CGA AGG ATG TCC	1026
Arg Asn Cys Thr Cys Lys Ser Gly Ala Val Glu Cys Arg Arg Met Ser	
295 300 305	
TGT CCC CCT CTC AAT TGC TCC CCA GAC TCC CTC CCA GTA CAC ATT GCT	1074
Cys Pro Pro Leu Asn Cys Ser Pro Asp Ser Leu Pro Val His Ile Ala	
310 315 320	
GGC CAG TGC TGT AAG GTC TGC CGA CCA AAA TGT ATC TAT GGA GGA AAA	1122
Gly Gln Cys Cys Lys Val Cys Arg Pro Lys Cys Ile Tyr Gly Gly Lys	
325 330 335 340	
GTT CTT GCA GAA GGC CAG CGG ATT TTA ACC AAG AGC TGT CGG GAA TGC	1170
Val Leu Ala Glu Gly Gln Arg Ile Leu Thr Lys Ser Cys Arg Glu Cys	
345 350 355	
CGA GGT GGA GTT TTA GTA AAA ATT ACA GAA ATG TGT CCT CCT TTG AAC	1218
Arg Gly Gly Val Leu Val Lys Ile Thr Glu Met Cys Pro Pro Leu Asn	
360 365 370	
TGC TCA GAA AAG GAT CAC ATT CTT CCT GAG AAT CAG TGC TGC CGT GTC	1266
Cys Ser Glu Lys Asp His Ile Leu Pro Glu Asn Gln Cys Cys Arg Val	
375 380 385	
TGT AGA GGT CAT AAC TTT TGT GCA GAA GGA CCT AAA TGT GGT GAA AAC	1314
Cys Arg Gly His Asn Phe Cys Ala Glu Gly Pro Lys Cys Gly Glu Asn	
390 395 400	
TCA GAG TGC AAA AAC TGG AAT ACA AAA GCT ACT TGT GAG TGC AAG AGT	1362

Ser	Glu	Cys	Lys	Asn	Trp	Asn	Thr	Lys	Ala	Thr	Cys	Glu	Cys	Lys	Ser	
405					410					415					420	
GGT	TAC	ATC	TCT	GTC	CAG	GGA	GAC	TCT	GCC	TAC	TGT	GAA	GAT	ATT	GAT	1410
Gly	Tyr	Ile	Ser	Val	Gln	Gly	Asp	Ser	Ala	Tyr	Cys	Glu	Asp	Ile	Asp	
				425					430					435		
GAG	TGT	GCA	GCT	AAG	ATG	CAT	TAC	TGT	CAT	GCC	AAT	ACT	GTG	TGT	GTC	1458
Glu	Cys	Ala	Ala	Lys	Met	His	Tyr	Cys	His	Ala	Asn	Thr	Val	Cys	Val	
				440				445					450			
AAC	CTT	CCT	GGG	TTA	TAT	CGC	TGT	GAC	TGT	GTC	CCA	GGA	TAC	ATT	CGT	1506
Asn	Leu	Pro	Gly	Leu	Tyr	Arg	Cys	Asp	Cys	Val	Pro	Gly	Tyr	Ile	Arg	
		455					460					465				
GTG	GAT	GAC	TTC	TCT	TGT	ACA	GAA	CAC	GAT	GAA	TGT	GGC	AGC	GGC	CAG	1554
Val	Asp	Asp	Phe	Ser	Cys	Thr	Glu	His	Asp	Glu	Cys	Gly	Ser	Gly	Gln	
	470					475					480					
CAC	AAC	TGT	GAT	GAG	AAT	GCC	ATC	TGC	ACC	AAC	ACT	GTC	CAG	GGA	CAC	1602
His	Asn	Cys	Asp	Glu	Asn	Ala	Ile	Cys	Thr	Asn	Thr	Val	Gln	Gly	His	
485					490					495					500	
AGC	TGC	ACC	TGC	AAA	CCG	GGC	TAC	GTG	GGG	AAC	GGG	ACC	ATC	TGC	AGA	1650
Ser	Cys	Thr	Cys	Lys	Pro	Gly	Tyr	Val	Gly	Asn	Gly	Thr	Ile	Cys	Arg	
				505					510					515		
GCT	TTC	TGT	GAA	GAG	GGC	TGC	AGA	TAC	GGT	GGA	ACG	TGT	GTG	GCT	CCC	1698
Ala	Phe	Cys	Glu	Glu	Gly	Cys	Arg	Tyr	Gly	Gly	Thr	Cys	Val	Ala	Pro	
			520					525					530			
AAC	AAA	TGT	GTC	TGT	CCA	TCT	GGA	TTC	ACA	GGA	AGC	CAC	TGC	GAG	AAA	1746
Asn	Lys	Cys	Val	Cys	Pro	Ser	Gly	Phe	Thr	Gly	Ser	His	Cys	Glu	Lys	
		535					540					545				
GAT	ATT	GAT	GAA	TGT	TCA	GAG	GGA	ATC	ATT	GAG	TGC	CAC	AAC	CAT	TCC	1794
Asp	Ile	Asp	Glu	Cys	Ser	Glu	Gly	Ile	Ile	Glu	Cys	His	Asn	His	Ser	
	550					555					560					
CGC	TGC	GTT	AAC	CTG	CCA	GGG	TGG	TAC	CAC	TGT	GAG	TGC	AGA	AGC	GGT	1842
Arg	Cys	Val	Asn	Leu	Pro	Gly	Trp	Tyr	His	Cys	Glu	Cys	Arg	Ser	Gly	
565				570						575					580	
TTC	CAT	GAC	GAT	GGG	ACC	TAT	TCA	CTG	TCC	GGG	GAG	TCC	TGT	ATT	GAC	1890
Phe	His	Asp	Asp	Gly	Thr	Tyr	Ser	Leu	Ser	Gly	Glu	Ser	Cys	Ile	Asp	
				585					590					595		
ATT	GAT	GAA	TGT	GCC	TTA	AGA	ACT	CAC	ACC	TGT	TGG	AAC	GAT	TCT	GCC	1938
Ile	Asp	Glu	Cys	Ala	Leu	Arg	Thr	His	Thr	Cys	Trp	Asn	Asp	Ser	Ala	

600	605	610	
TGC ATC AAC CTG GCA GGG GGT TTT GAC TGT CTC TGC CCC TCT GGG CCC Cys Ile Asn Leu Ala Gly Gly Phe Asp Cys Leu Cys Pro Ser Gly Pro 615 620 625			1986
TCC TGC TCT GGT GAC TGT CCT CAT GAA GGG GGG CTG AAG CAC AAT GGC Ser Cys Ser Gly Asp Cys Pro His Glu Gly Gly Leu Lys His Asn Gly 630 635 640			2034
CAG GTG TGG ACC TTG AAA GAA GAC AGG TGT TCT GTC TGC TCC TGC AAG Gln Val Trp Thr Leu Lys Glu Asp Arg Cys Ser Val Cys Ser Cys Lys 645 650 655 660			2082
GAT GGC AAG ATA TTC TGC CGA CGG ACA GCT TGT GAT TGC CAG AAT CCA Asp Gly Lys Ile Phe Cys Arg Arg Thr Ala Cys Asp Cys Gln Asn Pro 665 670 675			2130
AGT GCT GAC CTA TTC TGT TGC CCA GAA TGT GAC ACC AGA GTC ACA AGT Ser Ala Asp Leu Phe Cys Cys Pro Glu Cys Asp Thr Arg Val Thr Ser 680 685 690			2178
CAA TGT TTA GAC CAA AAT GGT CAC AAG CTG TAT CGA AGT GGA GAC AAT Gln Cys Leu Asp Gln Asn Gly His Lys Leu Tyr Arg Ser Gly Asp Asn 695 700 705			2226
TGG ACC CAT AGC TGT CAG CAG TGT CGG TGT CTG GAA GGA GAG GTA GAT Trp Thr His Ser Cys Gln Gln Cys Arg Cys Leu Glu Gly Glu Val Asp 710 715 720			2274
TGC TGG CCA CTC ACT TGC CCC AAC TTG AGC TGT GAG TAT ACA GCT ATC Cys Trp Pro Leu Thr Cys Pro Asn Leu Ser Cys Glu Tyr Thr Ala Ile 725 730 735 740			2322
TTA GAA GGG GAA TGT TGT CCC CGC TGT GTC AGT GAC CCC TGC CTA GCT Leu Glu Gly Glu Cys Cys Pro Arg Cys Val Ser Asp Pro Cys Leu Ala 745 750 755			2370
GAT AAC ATC ACC TAT GAC ATC AGA AAA ACT TGC CTG GAC AGC TAT GGT Asp Asn Ile Thr Tyr Asp Ile Arg Lys Thr Cys Leu Asp Ser Tyr Gly 760 765 770			2418
GTT TCA CGG CTT AGT GGC TCA GTG TGG ACG ATG GCT GGA TCT CCC TGC Val Ser Arg Leu Ser Gly Ser Val Trp Thr Met Ala Gly Ser Pro Cys 775 780 785			2466
ACA ACC TGT AAA TGC AAG AAT GGA AGA GTC TGT TGT TCT GTG GAT TTT Thr Thr Cys Lys Cys Lys Asn Gly Arg Val Cys Cys Ser Val Asp Phe 790 795 800			2514

GAG TGT CTT CAA AAT AAT TGAAGTATTT ACAGTGGACT CAACGCAGAA	2562
Glu Cys Leu Gln Asn Asn	
805 810	
GAATGGACGA AATGACCATC CAACGTGATT AAGGATAGGA ATCGGTAGTT TGGTTTTTTTT	2622
GTTTGTTTTG TTTTTTTAAC CACAGATAAT TGCCAAAGTT TCCACCTGAG GACGGTGTTT	2682
CGGAGGTTGC CTTTTGGACC TACCACTTTG CTCATTCTTG CTAACCTAGT CTAGGTGACC	2742
TACAGTCCCG TGCATTTAAG TCAATGGTTG TTAAAAGAAG TTTCCCGTGT TGTAATCAT	2802
GTTTCCCTTA TCAGATCATT TGCAAATACA TTAAATGAT CTCATGGTAA ATGGTTGATG	2862
TATTTTTTGG GTTATTTTG TGTACTAACC ATAATAGAGA GAGACTCAGC TCCTTTTATT	2922
TATTTGTG TGTTATGGAT CAAATCTAA AATAAAGTTG CCTGTTGTGA CTTTT	2977

[0348]

SEQ ID NO:37

SEQUENCE CHARACTERISTICS:

LENGTH: 816 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met	Glu	Ser	Arg	Val	Leu	Leu	Arg	Thr	Phe	Cys	Leu	Ile	Phe	Gly	Leu	
1					5				10					15		
Gly	Ala	Val	Trp	Gly	Leu	Gly	Val	Asp	Pro	Ser	Leu	Gln	Ile	Asp	Val	
			20					25						30		
Leu	Thr	Glu	Leu	Glu	Leu	Gly	Glu	Ser	Thr	Thr	Gly	Val	Arg	Gln	Val	
		35					40						45			
Pro	Gly	Leu	His	Asn	Gly	Thr	Lys	Ala	Phe	Leu	Phe	Gln	Asp	Thr	Pro	
	50					55					60					
Arg	Ser	Ile	Lys	Ala	Ser	Thr	Ala	Thr	Ala	Glu	Gln	Phe	Phe	Gln	Lys	
65					70					75					80	

Leu	Arg	Asn	Lys	His	Glu	Phe	Thr	Ile	Leu	Val	Thr	Leu	Lys	Gln	Thr
				85					90					95	
His	Leu	Asn	Ser	Gly	Val	Ile	Leu	Ser	Ile	His	His	Leu	Asp	His	Arg
			100					105					110		
Tyr	Leu	Glu	Leu	Glu	Ser	Ser	Gly	His	Arg	Asn	Glu	Val	Arg	Leu	His
		115					120					125			
Tyr	Arg	Ser	Gly	Ser	His	Arg	Pro	His	Thr	Glu	Val	Phe	Pro	Tyr	Ile
	130					135					140				
Leu	Ala	Asp	Asp	Lys	Trp	His	Lys	Leu	Ser	Leu	Ala	Ile	Ser	Ala	Ser
145					150					155					160
His	Leu	Ile	Leu	His	Ile	Asp	Cys	Asn	Lys	Ile	Tyr	Glu	Arg	Val	Val
				165					170					175	
Glu	Lys	Pro	Ser	Thr	Asp	Leu	Pro	Leu	Gly	Thr	Thr	Phe	Trp	Leu	Gly
			180					185					190		
Gln	Arg	Asn	Asn	Ala	His	Gly	Tyr	Phe	Lys	Gly	Ile	Met	Gln	Asp	Val
		195					200					205			
Gln	Leu	Leu	Val	Met	Pro	Gln	Gly	Phe	Ile	Ala	Gln	Cys	Pro	Asp	Leu
	210					215					220				
Asn	Arg	Thr	Cys	Pro	Thr	Cys	Asn	Asp	Phe	His	Gly	Leu	Val	Gln	Lys
225					230					235					240
Ile	Met	Glu	Leu	Gln	Asp	Ile	Leu	Ala	Lys	Thr	Ser	Ala	Lys	Leu	Ser
				245					250					255	
Arg	Ala	Glu	Gln	Arg	Met	Asn	Arg	Leu	Asp	Gln	Cys	Tyr	Cys	Glu	Arg
			260					265					270		
Thr	Cys	Thr	Met	Lys	Gly	Thr	Thr	Tyr	Arg	Glu	Phe	Glu	Ser	Trp	Ile
		275					280					285			
Asp	Gly	Cys	Lys	Asn	Cys	Thr	Cys	Leu	Asn	Gly	Thr	Ile	Gln	Cys	Glu
	290					295					300				
Thr	Leu	Ile	Cys	Pro	Asn	Pro	Asp	Cys	Pro	Leu	Lys	Ser	Ala	Leu	Ala
305					310					315					320
Tyr	Val	Asp	Gly	Lys	Cys	Cys	Lys	Glu	Cys	Lys	Ser	Ile	Cys	Gln	Phe
				325					330					335	
Gln	Gly	Arg	Thr	Tyr	Phe	Glu	Gly	Glu	Arg	Asn	Thr	Val	Tyr	Ser	Ser

340	345	350
Ser Gly Val Cys Val Leu Tyr Glu Cys Lys Asp Gln Thr Met Lys Leu		
355	360	365
Val Glu Ser Ser Gly Cys Pro Ala Leu Asp Cys Pro Glu Ser His Gln		
370	375	380
Ile Thr Leu Ser His Ser Cys Cys Lys Val Cys Lys Gly Tyr Asp Phe		
385	390	395
Cys Ser Glu Arg His Asn Cys Met Glu Asn Ser Ile Cys Arg Asn Leu		
405	410	415
Asn Asp Arg Ala Val Cys Ser Cys Arg Asp Gly Phe Arg Ala Leu Arg		
420	425	430
Glu Asp Asn Ala Tyr Cys Glu Asp Ile Asp Glu Cys Ala Glu Gly Arg		
435	440	445
His Tyr Cys Arg Glu Asn Thr Met Cys Val Asn Thr Pro Gly Ser Phe		
450	455	460
Met Cys Ile Cys Lys Thr Gly Tyr Ile Arg Ile Asp Asp Tyr Ser Cys		
465	470	475
Thr Glu His Asp Glu Cys Ile Thr Asn Gln His Asn Cys Asp Glu Asn		
485	490	495
Ala Leu Cys Phe Asn Thr Val Gly Gly His Asn Cys Val Cys Lys Pro		
500	505	510
Gly Tyr Thr Gly Asn Gly Thr Thr Cys Lys Ala Phe Cys Lys Asp Gly		
515	520	525
Cys Arg Asn Gly Gly Ala Cys Ile Ala Ala Asn Val Cys Ala Cys Pro		
530	535	540
Gln Gly Phe Thr Gly Pro Ser Cys Glu Thr Asp Ile Asp Glu Cys Ser		
545	550	555
Asp Gly Phe Val Gln Cys Asp Ser Arg Ala Asn Cys Ile Asn Leu Pro		
565	570	575
Gly Trp Tyr His Cys Glu Cys Arg Asp Gly Tyr His Asp Asn Gly Met		
580	585	590
Phe Ser Pro Ser Gly Glu Ser Cys Glu Asp Ile Asp Glu Cys Gly Thr		
595	600	605

Gly	Arg	His	Ser	Cys	Ala	Asn	Asp	Thr	Ile	Cys	Phe	Asn	Leu	Asp	Gly
610						615					620				
Gly	Tyr	Asp	Cys	Arg	Cys	Pro	His	Gly	Lys	Asn	Cys	Thr	Gly	Asp	Cys
625					630				635						640
Ile	His	Asp	Gly	Lys	Val	Lys	His	Asn	Gly	Gln	Ile	Trp	Val	Leu	Glu
				645					650					655	
Asn	Asp	Arg	Cys	Ser	Val	Cys	Ser	Cys	Gln	Asn	Gly	Phe	Val	Met	Cys
			660					665					670		
Arg	Arg	Met	Val	Cys	Asp	Cys	Glu	Asn	Pro	Thr	Val	Asp	Leu	Phe	Cys
		675					680					685			
Cys	Pro	Glu	Cys	Asp	Pro	Arg	Leu	Ser	Ser	Gln	Cys	Leu	His	Gln	Asn
	690					695					700				
Gly	Glu	Thr	Leu	Tyr	Asn	Ser	Gly	Asp	Thr	Trp	Val	Gln	Asn	Cys	Gln
705					710					715					720
Gln	Cys	Arg	Cys	Leu	Gln	Gly	Glu	Val	Asp	Cys	Trp	Pro	Leu	Pro	Cys
				725					730					735	
Pro	Asp	Val	Glu	Cys	Glu	Phe	Ser	Ile	Leu	Pro	Glu	Asn	Glu	Cys	Cys
			740					745					750		
Pro	Arg	Cys	Val	Thr	Asp	Pro	Cys	Gln	Ala	Asp	Thr	Ile	Arg	Asn	Asp
		755					760					765			
Ile	Thr	Lys	Thr	Cys	Leu	Asp	Glu	Met	Asn	Val	Val	Arg	Phe	Thr	Gly
	770					775					780				
Ser	Ser	Trp	Ile	Lys	His	Gly	Thr	Glu	Cys	Thr	Leu	Cys	Gln	Cys	Lys
785					790					795					800
Asn	Gly	His	Ile	Cys	Cys	Ser	Val	Asp	Pro	Gln	Cys	Leu	Gln	Glu	Leu
				805					810					815	

[0349]

SEQ ID NO:38

SEQUENCE CHARACTERISTICS:

LENGTH: 2448 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGGAGTCTC GGGTCTTACT GAGAACATTC TGTTTGATCT TCGGTCTCGG AGCAGTTTGG	60
GGGCTTGGTG TGGACCCCTC CCTACAGATT GACGCTTAA CAGAGTTAGA ACTTGGGGAG	120
TCCACGACCG GAGTGCGTCA GGTCCCAGGG CTGCATAATG GGACGAAAGC CTTTCTCTTT	180
CAAGATACTC CCAGAAGCAT AAAAGCATCC ACTGCTACAG CTGAACAGTT TTTTCAGAAG	240
CTGAGAAATA AACATGAATT TACTATTTTG GTGACCCTAA AACAGACCCA CTTAAATTCA	300
GGAGTTATTC TCTCAATTCA CCACTTGGAT CACAGGTACC TGGAAGTGA AAGTAGTGGC	360
CATCGGAATG AAGTCAGACT GCATTACCGC TCAGGCAGTC ACCGCCCTCA CACAGAAGTG	420
TTTCCCTTACA TTTTGGCTGA TGACAAGTGG CACAAGCTCT CCTTAGCCAT CAGTGCTTCC	480
CATTTGATTT TACACATTGA CTGCAATAAA ATTTATGAAA GGGTAGTAGA AAAGCCCTCC	540
ACAGACTTGC CTCTAGGCAC AACATTTTGG CTAGGACAGA GAAATAATGC GCATGGATAT	600
TTTAAGGGTA TAATGCAAGA TGTCCAATTA CTTGTCATGC CCCAGGGATT TATTGCTCAG	660
TGCCCAGATC TTAATCGCAC CTGTCCAACT TGCAATGACT TCCATGGACT TGTGCAGAAA	720
ATCATGGAGC TACAGGATAT TTTAGCCAAA ACATCAGCCA AGCTGTCTCG AGCTGAACAG	780
CGAATGAATA GATTGGATCA GTGCTATTGT GAAAGGACTT GCACCATGAA GGAACCACC	840
TACCGAGAAT TTGAGTCCTG GATAGACGGC TGTAAGAACT GCACATGCCT GAATGGAACC	900
ATCCAGTGTG AACTCTAAT CTGCCCAAAT CCTGACTGCC CACTTAAGTC GGCTCTTGCG	960
TATGTGGATG GCAAATGCTG TAAGGAATGC AAATCGATAT GCCAATTTCA AGGACGAACC	1020
TACTTTGAAG GAGAAAGAAA TACAGTCTAT TCCTCTTCTG GAGTATGTGT TCTCTATGAG	1080
TGCAAGGACC AGACCATGAA ACTTGTGAG AGTTCAGGCT GTCCAGCTTT GGATTGTCCA	1140
GAGTCTCATC AGATAACCTT GTCTCACAGC TGTTGCAAAG TTTGTAAAGG TTATGACTTT	1200
TGTTCTGAAA GGCATAACTG CATGGAGAAT TCCATCTGCA GAAATCTGAA TGACAGGGCT	1260

GTTTGTAGCT	GTCGAGATGG	TTTTAGGGCT	CTTCGAGAGG	ATAATGCCTA	CTGTGAAGAC	1320
ATCGATGAGT	GTCGTAAGG	GCGCCATTAC	TGTCGTGAAA	ATACAATGTG	TGTCAACACC	1380
CCGGGTTCCT	TTATGTGCAT	CTGCAAAACT	GGATACATCA	GAATTGATGA	TTATTTCATGT	1440
ACAGAACATG	ATGAGTGTAT	CACAAATCAG	CACAACTGTG	ATGAAAATGC	TTTATGCTTC	1500
AACACTGTTG	GAGGACACAA	CTGTGTTTGC	AAGCCGGGCT	ATACAGGGAA	TGGAACGACA	1560
TGCAAAGCAT	TTTGCAAAGA	TGGCTGTAGG	AATGGAGGAG	CCTGTATTGC	CGCTAATGTG	1620
TGTGCCTGCC	CACAAGGCTT	CACTGGACCC	AGCTGTGAAA	CGGACATTGA	TGAATGCTCT	1680
GATGGTTTTG	TTCAATGTGA	CAGTCGTGCT	AATTGCATTA	ACCTGCCTGG	ATGGTACCAC	1740
TGTGAGTGCA	GAGATGGCTA	CCATGACAAT	GGGATGTTTT	CACCAAGTGG	AGAATCGTGT	1800
GAAGATATTG	ATGAGTGTGG	GACCGGGAGG	CACAGCTGTG	CCAATGATAC	CATTTGCTTC	1860
AAFTTGGATG	GCGGATATGA	TTGTTCGATG	CCTCATGGAA	AGAATTGCAC	AGGGGACTGC	1920
ATCCATGATG	GAAAAGTTAA	GCACAATGGT	CAGATTTGGG	TGTTGGAAAA	TGACAGGTGC	1980
TCTGTGTGCT	CATGTCAGAA	TGGATTCTGT	ATGTGTTCGAC	GGATGGTCTG	TGACTGTGAG	2040
AATCCACAG	TTGATCTTTT	TTGCTGCCCT	GAATGTGACC	CAAGGCTTAG	TAGTCAGTGC	2100
CTCCATCAAA	ATGGGGAAAC	TTTGTATAAC	AGTGGTGACA	CCTGGGTCCA	GAATTGTCAA	2160
CAGTGCCGCT	GCTTGCAAGG	GGAAGTTGAT	TGTTGGCCCC	TGCCTTGCCC	AGATGTGGAG	2220
TGTGAATTCA	GCATTCTCCC	AGAGAATGAG	TGCTGCCCCG	GCTGTGTCAC	AGACCCTTGC	2280
CAGGCTGACA	CCATCCGCAA	TGACATCACC	AAGACTTGCC	TGGACGAAAT	GAATGTGGTT	2340
CGCTTCACCG	GGTCCTCTTG	GATCAAACAT	GGCACTGAGT	GTACTCTCTG	CCAGTGCAAG	2400
AATGGCCACA	TCTGTTGCTC	AGTGGATCCA	CAGTGCCCTC	AGGAACTG		2448

[0350]

SEQ ID NO:39

SEQUENCE CHARACTERISTICS:

LENGTH: 3198 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-093E05

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 97..2544

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

TTGGGAGGAG CAGTCTCTCC GCTCGTCTCC CGGAGCTTTC TCCATTGTCT CTGCCTTTAC	60
AACAGAGGGA GACGATGGAC TGAGCTGATC CGCACC ATG GAG TCT CGG GTC TTA	114
Met Glu Ser Arg Val Leu	
1 5	
CTG AGA ACA TTC TGT TTG ATC TTC GGT CTC GGA GCA GTT TGG GGG CTT	162
Leu Arg Thr Phe Cys Leu Ile Phe Gly Leu Gly Ala Val Trp Gly Leu	
10 15 20	
GGT GTG GAC CCT TCC CTA CAG ATT GAC GTC TTA ACA GAG TTA GAA CTT	210
Gly Val Asp Pro Ser Leu Gln Ile Asp Val Leu Thr Glu Leu Glu Leu	
25 30 35	
GGG GAG TCC ACG ACC GGA GTG CGT CAG GTC CCG GGG CTG CAT AAT GGG	258
Gly Glu Ser Thr Thr Gly Val Arg Gln Val Pro Gly Leu His Asn Gly	
40 45 50	
ACG AAA GCC TTT CTC TTT CAA GAT ACT CCC AGA AGC ATA AAA GCA TCC	306
Thr Lys Ala Phe Leu Phe Gln Asp Thr Pro Arg Ser Ile Lys Ala Ser	
55 60 65 70	
ACT GCT ACA GCT GAA CAG TTT TTT CAG AAG CTG AGA AAT AAA CAT GAA	354
Thr Ala Thr Ala Glu Gln Phe Phe Gln Lys Leu Arg Asn Lys His Glu	
75 80 85	

TTT ACT ATT TTG GTG ACC CTA AAA CAG ACC CAC TTA AAT TCA GGA GTT	402
Phe Thr Ile Leu Val Thr Leu Lys Gln Thr His Leu Asn Ser Gly Val	
90 95 100	
ATT CTC TCA ATT CAC CAC TTG GAT CAC AGG TAC CTG GAA CTG GAA AGT	450
Ile Leu Ser Ile His His Leu Asp His Arg Tyr Leu Glu Leu Glu Ser	
105 110 115	
AGT GGC CAT CGG AAT GAA GTC AGA CTG CAT TAC CGC TCA GGC AGT CAC	498
Ser Gly His Arg Asn Glu Val Arg Leu His Tyr Arg Ser Gly Ser His	
120 125 130	
CGC CCT CAC ACA GAA GTG TTT CCT TAC ATT TTG GCT GAT GAC AAG TGG	546
Arg Pro His Thr Glu Val Phe Pro Tyr Ile Leu Ala Asp Asp Lys Trp	
135 140 145 150	
CAC AAG CTC TCC TTA GCC ATC AGT GCT TCC CAT TTG ATT TTA CAC ATT	594
His Lys Leu Ser Leu Ala Ile Ser Ala Ser His Leu Ile Leu His Ile	
155 160 165	
GAC TGC AAT AAA ATT TAT GAA AGG GTA GTA GAA AAG CCC TCC ACA GAC	642
Asp Cys Asn Lys Ile Tyr Glu Arg Val Val Glu Lys Pro Ser Thr Asp	
170 175 180	
TTG CCT CTA GGC ACA ACA TTT TGG CTA GGA CAG AGA AAT AAT GCG CAT	690
Leu Pro Leu Gly Thr Thr Phe Trp Leu Gly Gln Arg Asn Asn Ala His	
185 190 195	
GGA TAT TTT AAG GGT ATA ATG CAA GAT GTC CAA TTA CTT GTC ATG CCC	738
Gly Tyr Phe Lys Gly Ile Met Gln Asp Val Gln Leu Leu Val Met Pro	
200 205 210	
CAG GGA TTT ATT GCT CAG TGC CCA GAT CTT AAT CGC ACC TGT CCA ACT	786
Gln Gly Phe Ile Ala Gln Cys Pro Asp Leu Asn Arg Thr Cys Pro Thr	
215 220 225 230	
TGC AAT GAC TTC CAT GGA CTT GTG CAG AAA ATC ATG GAG CTA CAG GAT	834
Cys Asn Asp Phe His Gly Leu Val Gln Lys Ile Met Glu Leu Gln Asp	
235 240 245	
ATT TTA GCC AAA ACA TCA GCC AAG CTG TCT CGA GCT GAA CAG CGA ATG	882
Ile Leu Ala Lys Thr Ser Ala Lys Leu Ser Arg Ala Glu Gln Arg Met	
250 255 260	
AAT AGA TTG GAT CAG TGC TAT TGT GAA AGG ACT TGC ACC ATG AAG GGA	930
Asn Arg Leu Asp Gln Cys Tyr Cys Glu Arg Thr Cys Thr Met Lys Gly	
265 270 275	
ACC ACC TAC CGA GAA TTT GAG TCC TGG ATA GAC GGC TGT AAG AAC TGC	978

Thr	Thr	Tyr	Arg	Glu	Phe	Glu	Ser	Trp	Ile	Asp	Gly	Cys	Lys	Asn	Cys	
280						285					290					
ACA	TGC	CTG	AAT	GGA	ACC	ATC	CAG	TGT	GAA	ACT	CTA	ATC	TGC	CCA	AAT	1026
Thr	Cys	Leu	Asn	Gly	Thr	Ile	Gln	Cys	Glu	Thr	Leu	Ile	Cys	Pro	Asn	
295					300				305						310	
CCT	GAC	TGC	CCA	CTT	AAG	TCG	GCT	CTT	GCG	TAT	GTG	GAT	GGC	AAA	TGC	1074
Pro	Asp	Cys	Pro	Leu	Lys	Ser	Ala	Leu	Ala	Tyr	Val	Asp	Gly	Lys	Cys	
				315					320					325		
TGT	AAG	GAA	TGC	AAA	TCG	ATA	TGC	CAA	TTT	CAA	GGA	CGA	ACC	TAC	TTT	1122
Cys	Lys	Glu	Cys	Lys	Ser	Ile	Cys	Gln	Phe	Gln	Gly	Arg	Thr	Tyr	Phe	
			330					335					340			
GAA	GGA	GAA	AGA	AAT	ACA	GTC	TAT	TCC	TCT	TCT	GGA	GTA	TGT	GTT	CTC	1170
Glu	Gly	Glu	Arg	Asn	Thr	Val	Tyr	Ser	Ser	Ser	Gly	Val	Cys	Val	Leu	
		345					350					355				
TAT	GAG	TGC	AAG	GAC	CAG	ACC	ATG	AAA	CTT	GTT	GAG	AGT	TCA	GGC	TGT	1218
Tyr	Glu	Cys	Lys	Asp	Gln	Thr	Met	Lys	Leu	Val	Glu	Ser	Ser	Gly	Cys	
	360					365					370					
CCA	GCT	TTG	GAT	TGT	CCA	GAG	TCT	CAT	CAG	ATA	ACC	TTG	TCT	CAC	AGC	1266
Pro	Ala	Leu	Asp	Cys	Pro	Glu	Ser	His	Gln	Ile	Thr	Leu	Ser	His	Ser	
375					380					385					390	
TGT	TGC	AAA	GTT	TGT	AAA	GGT	TAT	GAC	TTT	TGT	TCT	GAA	AGG	CAT	AAC	1314
Cys	Cys	Lys	Val	Cys	Lys	Gly	Tyr	Asp	Phe	Cys	Ser	Glu	Arg	His	Asn	
				395					400					405		
TGC	ATG	GAG	AAT	TCC	ATC	TGC	AGA	AAT	CTG	AAT	GAC	AGG	GCT	GTT	TGT	1362
Cys	Met	Glu	Asn	Ser	Ile	Cys	Arg	Asn	Leu	Asn	Asp	Arg	Ala	Val	Cys	
			410					415					420			
AGC	TGT	CGA	GAT	GGT	TTT	AGG	GCT	CTT	CGA	GAG	GAT	AAT	GCC	TAC	TGT	1410
Ser	Cys	Arg	Asp	Gly	Phe	Arg	Ala	Leu	Arg	Glu	Asp	Asn	Ala	Tyr	Cys	
		425					430					435				
GAA	GAC	ATC	GAT	GAG	TGT	GCT	GAA	GGG	CGC	CAT	TAC	TGT	CGT	GAA	AAT	1458
Glu	Asp	Ile	Asp	Glu	Cys	Ala	Glu	Gly	Arg	His	Tyr	Cys	Arg	Glu	Asn	
	440					445					450					
ACA	ATG	TGT	GTC	AAC	ACC	CCG	GGT	TCT	TTT	ATG	TGC	ATC	TGC	AAA	ACT	1506
Thr	Met	Cys	Val	Asn	Thr	Pro	Gly	Ser	Phe	Met	Cys	Ile	Cys	Lys	Thr	
455					460					465					470	
GGA	TAC	ATC	AGA	ATT	GAT	GAT	TAT	TCA	TGT	ACA	GAA	CAT	GAT	GAG	TGT	1554
Gly	Tyr	Ile	Arg	Ile	Asp	Asp	Tyr	Ser	Cys	Thr	Glu	His	Asp	Glu	Cys	

475								480								485	
ATC	ACA	AAT	CAG	CAC	AAC	TGT	GAT	GAA	AAT	GCT	TTA	TGC	TTC	AAC	ACT	1602	
Ile	Thr	Asn	Gln	His	Asn	Cys	Asp	Glu	Asn	Ala	Leu	Cys	Phe	Asn	Thr		
490								495								500	
GTT	GGA	GGA	CAC	AAC	TGT	GTT	TGC	AAG	CCG	GGC	TAT	ACA	GGG	AAT	GGA	1650	
Val	Gly	Gly	His	Asn	Cys	Val	Cys	Lys	Pro	Gly	Tyr	Thr	Gly	Asn	Gly		
505								510								515	
ACG	ACA	TGC	AAA	GCA	TTT	TGC	AAA	GAT	GGC	TGT	AGG	AAT	GGA	GGA	GCC	1698	
Thr	Thr	Cys	Lys	Ala	Phe	Cys	Lys	Asp	Gly	Cys	Arg	Asn	Gly	Gly	Ala		
520								525								530	
TGT	ATT	GCC	GCT	AAT	GTG	TGT	GCC	TGC	CCA	CAA	GGC	TTC	ACT	GGA	CCC	1746	
Cys	Ile	Ala	Ala	Asn	Val	Cys	Ala	Cys	Pro	Gln	Gly	Phe	Thr	Gly	Pro		
535								540								545	
AGC	TGT	GAA	ACG	GAC	ATT	GAT	GAA	TGC	TCT	GAT	GGT	TTT	GTT	CAA	TGT	1794	
Ser	Cys	Glu	Thr	Asp	Ile	Asp	Glu	Cys	Ser	Asp	Gly	Phe	Val	Gln	Cys		
555								560								565	
GAC	AGT	CGT	GCT	AAT	TGC	ATT	AAC	CTG	CCT	GGA	TGG	TAC	CAC	TGT	GAG	1842	
Asp	Ser	Arg	Ala	Asn	Cys	Ile	Asn	Leu	Pro	Gly	Trp	Tyr	His	Cys	Glu		
570								575								580	
TGC	AGA	GAT	GGC	TAC	CAT	GAC	AAT	GGG	ATG	TTT	TCA	CCA	AGT	GGA	GAA	1890	
Cys	Arg	Asp	Gly	Tyr	His	Asp	Asn	Gly	Met	Phe	Ser	Pro	Ser	Gly	Glu		
585								590								595	
TCG	TGT	GAA	GAT	ATT	GAT	GAG	TGT	GGG	ACC	GGG	AGG	CAC	AGC	TGT	GCC	1938	
Ser	Cys	Glu	Asp	Ile	Asp	Glu	Cys	Gly	Thr	Gly	Arg	His	Ser	Cys	Ala		
600								605								610	
AAT	GAT	ACC	ATT	TGC	TTC	AAT	TTG	GAT	GGC	GGA	TAT	GAT	TGT	CGA	TGT	1986	
Asn	Asp	Thr	Ile	Cys	Phe	Asn	Leu	Asp	Gly	Gly	Tyr	Asp	Cys	Arg	Cys		
615								620								625	
CCT	CAT	GGA	AAG	AAT	TGC	ACA	GGG	GAC	TGC	ATC	CAT	GAT	GGA	AAA	GTT	2034	
Pro	His	Gly	Lys	Asn	Cys	Thr	Gly	Asp	Cys	Ile	His	Asp	Gly	Lys	Val		
635								640								645	
AAG	CAC	AAT	GGT	CAG	ATT	TGG	GTG	TTG	GAA	AAT	GAC	AGG	TGC	TCT	GTG	2082	
Lys	His	Asn	Gly	Gln	Ile	Trp	Val	Leu	Glu	Asn	Asp	Arg	Cys	Ser	Val		
650								655								660	
TGC	TCA	TGT	CAG	AAT	GGA	TTC	GTT	ATG	TGT	CGA	CGG	ATG	GTC	TGT	GAC	2130	
Cys	Ser	Cys	Gln	Asn	Gly	Phe	Val	Met	Cys	Arg	Arg	Met	Val	Cys	Asp		
665								670								675	

TGT GAG AAT CCC ACA GTT GAT CTT TTT TGC TGC CCT GAA TGT GAC CCA Cys Glu Asn Pro Thr Val Asp Leu Phe Cys Cys Pro Glu Cys Asp Pro 680 685 690	2178
AGG CTT AGT AGT CAG TGC CTC CAT CAA AAT GGG GAA ACT TTG TAT AAC Arg Leu Ser Ser Gln Cys Leu His Gln Asn Gly Glu Thr Leu Tyr Asn 695 700 705 710	2226
AGT GGT GAC ACC TGG GTC CAG AAT TGT CAA CAG TGC CGC TGC TTG CAA Ser Gly Asp Thr Trp Val Gln Asn Cys Gln Gln Cys Arg Cys Leu Gln 715 720 725	2274
GGG GAA GTT GAT TGT TGG CCC CTG CCT TGC CCA GAT GTG GAG TGT GAA Gly Glu Val Asp Cys Trp Pro Leu Pro Cys Pro Asp Val Glu Cys Glu 730 735 740	2322
TTC AGC ATT CTC CCA GAG AAT GAG TGC TGC CCG CGC TGT GTC ACA GAC Phe Ser Ile Leu Pro Glu Asn Glu Cys Cys Pro Arg Cys Val Thr Asp 745 750 755	2370
CCT TGC CAG GCT GAC ACC ATC CGC AAT GAC ATC ACC AAG ACT TGC CTG Pro Cys Gln Ala Asp Thr Ile Arg Asn Asp Ile Thr Lys Thr Cys Leu 760 765 770	2418
GAC GAA ATG AAT GTG GTT CGC TTC ACC GGG TCC TCT TGG ATC AAA CAT Asp Glu Met Asn Val Val Arg Phe Thr Gly Ser Ser Trp Ile Lys His 775 780 785 790	2466
GGC ACT GAG TGT ACT CTC TGC CAG TGC AAG AAT GGC CAC ATC TGT TGC Gly Thr Glu Cys Thr Leu Cys Gln Cys Lys Asn Gly His Ile Cys Cys 795 800 805	2514
TCA GTG GAT CCA CAG TGC CTT CAG GAA CTG TGAAGTTAAC TGTCTCATGG Ser Val Asp Pro Gln Cys Leu Gln Glu Leu 810 815	2564
GAGATTTCCTG TTAAAAGAAT GTTCTTTTCAT TAAAAGACCA AAAAGAAGTT AAAACTTAAA	2624
TTGGGTGATT TGTGGGCAGC TAAATGCAGC TTTGTTAATA GCTGAGTGAA CTTTCAATTA	2684
TGAAATTTGT GGAGCTTGAC AAAATCACAA AAGGAAAATT ACTGGGGCAA AATTAGACCT	2744
CAAGTCTGCC TCTACTGTGT CTCACATCAC CATGTAGAAG AATGGGCGTA CAGTATATAC	2804
CGTGACATCC TGAACCCCTGG ATAGAAAGCC TGAGCCCATT GGATCTGTGA AAGCCTCTAG	2864
CTTCACTGGT GCAGAAAATT TTCCTCTAGA TCAGAATCTT CAGAATCAGT TAGGTTCCCTC	2924
ACTGCAAGAA ATAAAATGTC AGGCAGTGAA TGAATTATAT TTTCAGAAGT AAAGCAAAGA	2984

AGCTATAACA TGTTATGTAC AGTACACTCT GAAAAGAAAT CTGAAACAAG TTATTGTAAT 3044
 GATAAAAATA ATGCACAGGC ATGGTTACTT AATATTTTCT AACAGGAAAA GTCATCCCTA 3104
 TTTCCTTGTT TTAGTGCACT TAATATTATT TGGTTGAATT TGTTTCAGTAT AAGCTCGTTC 3164
 TTGTGCAAAA TTAAATAAAT ATTICTCTTA CCTT 3198

[0351]

SEQ ID NO:40

SEQUENCE CHARACTERISTICS:

LENGTH: 499 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION:

Met	Glu	Leu	Ser	Glu	Pro	Val	Val	Glu	Asn	Gly	Glu	Val	Glu	Met	Ala	1	5	10	15
Leu	Glu	Glu	Ser	Trp	Glu	His	Ser	Lys	Glu	Val	Ser	Glu	Ala	Glu	Pro	20	25	30	
Gly	Gly	Gly	Ser	Ser	Gly	Asp	Ser	Gly	Pro	Pro	Glu	Glu	Ser	Gly	Gln	35	40	45	
Glu	Met	Met	Glu	Glu	Lys	Glu	Glu	Ile	Arg	Lys	Ser	Lys	Ser	Val	Ile	50	55	60	
Val	Pro	Ser	Gly	Ala	Pro	Lys	Lys	Glu	His	Val	Asn	Val	Val	Phe	Ile	65	70	75	80
Gly	His	Val	Asp	Ala	Gly	Lys	Ser	Thr	Ile	Gly	Gly	Gln	Ile	Met	Phe	85	90	95	
Leu	Thr	Gly	Met	Ala	Asp	Lys	Arg	Thr	Leu	Glu	Lys	Tyr	Glu	Arg	Glu	100	105	110	
Ala	Glu	Glu	Lys	Asn	Arg	Glu	Thr	Trp	Tyr	Leu	Ser	Trp	Ala	Leu	Asp	115	120	125	

Thr Asn Gln Glu Glu Arg Asp Lys Gly Lys Thr Val Glu Val Gly Arg
130 135 140

Ala Tyr Phe Glu Thr Glu Arg Lys His Phe Thr Ile Leu Asp Ala Pro
145 150 155 160

Gly His Lys Ser Phe Val Pro Asn Met Ile Gly Gly Ala Ser Gln Ala
165 170 175

Asp Leu Ala Val Leu Val Ile Ser Ala Arg Lys Gly Glu Phe Glu Thr
180 185 190

Gly Phe Glu Lys Gly Gly Gln Thr Arg Glu His Ala Met Phe Gly Lys
195 200 205

Thr Ala Gly Val Lys His Leu Ile Val Leu Ile Asn Lys Met Asp Asp
210 215 220

Pro Thr Val Asn Trp Gly Ile Glu Arg Tyr Glu Glu Cys Lys Glu Lys
225 230 235 240

Leu Val Pro Phe Leu Lys Lys Val Gly Phe Ser Pro Lys Lys Asp Ile
245 250 255

His Phe Met Pro Cys Ser Gly Leu Thr Gly Ala Asn Ile Lys Glu Gln
260 265 270

Ser Asp Phe Cys Pro Trp Tyr Thr Gly Leu Pro Phe Ile Pro Tyr Leu
275 280 285

Asn Asn Leu Pro Asn Phe Asn Arg Ser Ile Asp Gly Pro Ile Arg Leu
290 295 300

Pro Ile Val Asp Lys Tyr Lys Asp Met Gly Thr Val Val Leu Gly Lys
305 310 315 320

Leu Glu Ser Gly Ser Ile Phe Lys Gly Gln Gln Leu Val Met Met Pro
325 330 335

Asn Lys His Asn Val Glu Val Leu Gly Ile Leu Ser Asp Asp Thr Glu
340 345 350

Thr Asp Phe Val Ala Pro Gly Glu Asn Leu Lys Ile Arg Leu Lys Gly
355 360 365

Ile Glu Glu Glu Glu Ile Leu Pro Glu Phe Ile Leu Cys Asp Pro Ser
370 375 380

Asn Leu Cys His Ser Gly Arg Thr Phe Asp Val Gln Ile Val Ile Ile

385		390		395		400
Glu His Lys Ser Ile	Ile Cys Pro Gly Tyr Asn Ala Val Leu His Ile					
	405		410		415	
His Thr Cys Ile Glu Glu Val Glu Ile Thr Ala Leu Ile Ser Leu Val						
	420		425		430	
Asp Lys Lys Ser Gly Glu Lys Ser Lys Thr Arg Pro Arg Phe Val Lys						
	435		440		445	
Gln Asp Gln Val Cys Ile Ala Arg Leu Arg Thr Ala Gly Thr Ile Cys						
	450		455		460	
Leu Glu Thr Phe Lys Asp Phe Pro Gln Met Gly Arg Phe Thr Leu Arg						
	465		470		475	
Asp Glu Gly Lys Thr Ile Ala Ile Gly Lys Val Leu Lys Leu Val Pro						
	485		490		495	
Glu Lys Asp						

[0352]

SEQ ID NO:41

SEQUENCE CHARACTERISTICS:

LENGTH: 1497 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION:

ATGGAACCTTT CAGAACCTGT TGTAGAAAAT GGAGAGGTGG AAATGGCCCT AGAAGAATCA	60
TGGGAGCACA GTAAAGAAGT AAGTGAAGCC GAGCCTGGGG GTGGTTCCTC GGGAGATTCA	120
GGGCCCCCAG AAGAAAGTGG CCAGGAAATG ATGGAGGAAA AAGAGGAAAT AAGAAAATCC	180
AAATCTGTGA TCGTACCCTC AGGTGCACCT AAGAAAGAAC ACGTAAATGT AGTATTCATT	240

GGCCATGTAG	ACGCTGGCAA	GTCAACCATC	GGAGGACAGA	TAATGTTTTT	GA CTGGAATG	300
GCTGACAAAA	GAACACTGGA	GAAATATGAA	AGAGAAGCTG	AGGAAAAAAAA	CAGAGAAACC	360
TGGTATTTGT	CCTGGGCCCT	AGATACAAAT	CAGGAGGAAC	GAGACAAGGG	TAAAACAGTC	420
GAAGTGGGTC	GTGCCTATTT	TGAAACAGAA	AGGAAACATT	TCACAATTTT	AGATGCCCCCT	480
GGCCACAAGA	GTTTTGTCCC	AAATATGATT	GGTGGTGCTT	CTCAAGCTGA	TTTGGCTGTG	540
CTGGTCATCT	CTGCCAGGAA	AGGAGAGTTT	GAAACTGGAT	TTGAAAAAGG	TGGACAGACA	600
AGAGAACATG	CGATGTTTGG	CAAAACGGCA	GGAGTAAAC	ATTTAATAGT	GCTTATTAAT	660
AAGATGGATG	ATCCACAGT	AAATGGGGC	ATCGAGAGAT	ATGAAGAATG	TAAAGAAAAA	720
CTGGTGCCCT	TTTTGAAAAA	AGTAGGCTTT	AGTCCAAAAA	AGGACATTCA	CTTTATGCCC	780
TGCTCAGGAC	TGACCGGAGC	AAATATTAAA	GAGCAGTCAG	ATTTCTGCCC	TTGGTACACT	840
GGATTACCAT	TTATTCCGTA	TTTGAATAAC	TTGCCAAACT	TCAACAGATC	AATTGATGGA	900
CCAATAAGAC	TGCCAATTGT	GGATAAGTAC	AAAGATATGG	GCACTGTGGT	CCTGGGAAAG	960
CTGGAATCCG	GGTCCATTTT	TAAAGGCCAG	CAGCTCGTGA	TGATGCCAAA	CAAGCACAAAT	1020
GTAGAAGTTC	TTGGAATACT	TTCTGATGAT	ACTGAAACTG	ATTTTGTAGC	CCCAGGTGAA	1080
AACCTCAAAA	TCAGACTGAA	GGGAATTGAA	GAAGAAGAGA	TTCTTCCAGA	ATTCATACTT	1140
TGTGATCCTA	GTAACCTCTG	CCATTCTGGA	CGCACGTTTG	ATGTTCAGAT	AGTGATTATT	1200
GAGCACAAAT	CCATCATCTG	CCCAGGTTAT	AATGCGGTGC	TGCACATTCA	TACTTGTATT	1260
GAGGAAGTTG	AGATAACAGC	GTTAATCTCC	TTGGTAGACA	AAAAATCAGG	GGAAAAAAGT	1320
AAGACACGAC	CCCGCTTCGT	GAAACAAGAT	CAAGTATGCA	TTGCTCGTTT	AAGGACAGCA	1380
GGAACCATCT	GCCTCGAGAC	GTTCAAAGAT	TTTCCTCAGA	TGGGTCGTTT	TACTTTAAGA	1440
GATGAGGGTA	AGACCATTCG	AATTGGAAAA	GTTCTGAAAT	TGGTCCCAGA	GAAGGAC	1497

[0353]

SEQ ID NO:42

SEQUENCE CHARACTERISTICS:

LENGTH: 2057 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SOURCE:

LIBRARY: Human fetal brain cDNA library

CLONE: GEN-077A09

FEATURES OF THE SEQUENCE:

NAME/KEY: CDS

LOCATION: 144..1640

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION:

TCCCGGCCGG CTCCGGCAGC AACGATGAAG CCTGCACCGG CGCGGGATAC CCTCAAGGTA	60
AAAGGATGGG ACGGGGGGCA CCTGTGGAAC CTTCCCGAGA GGAACCGTTA GTGTCCGCTTG	120
AAGGTTCCAA TTCAGCCGTT ACC ATG GAA CTT TCA GAA CCT GTT GTA GAA	170
Met Glu Leu Ser Glu Pro Val Val Glu	
1 5	
AAT GGA GAG GTG GAA ATG GCC CTA GAA GAA TCA TGG GAG CAC AGT AAA	218
Asn Gly Glu Val Glu Met Ala Leu Glu Glu Ser Trp Glu His Ser Lys	
10 15 20 25	
GAA GTA AGT GAA GCC GAG CCT GGG GGT GGT TCC TCG GGA GAT TCA GGG	266
Glu Val Ser Glu Ala Glu Pro Gly Gly Gly Ser Ser Gly Asp Ser Gly	
30 35 40	
CCC CCA GAA GAA AGT GGC CAG GAA ATG ATG GAG GAA AAA GAG GAA ATA	314
Pro Pro Glu Glu Ser Gly Gln Glu Met Met Glu Glu Lys Glu Glu Ile	
45 50 55	
AGA AAA TCC AAA TCT GTG ATC GTA CCC TCA GGT GCA CCT AAG AAA GAA	362
Arg Lys Ser Lys Ser Val Ile Val Pro Ser Gly Ala Pro Lys Lys Glu	
60 65 70	

CAC	GTA	AAT	GTA	GTA	TTC	ATT	GGC	CAT	GTA	GAC	GCT	GGC	AAG	TCA	ACC	410
His	Val	Asn	Val	Val	Phe	Ile	Gly	His	Val	Asp	Ala	Gly	Lys	Ser	Thr	
	75					80					85					
ATC	GGA	GGA	CAG	ATA	ATG	TTT	TTG	ACT	GGA	ATG	GCT	GAC	AAA	AGA	ACA	458
Ile	Gly	Gly	Gln	Ile	Met	Phe	Leu	Thr	Gly	Met	Ala	Asp	Lys	Arg	Thr	
	90				95					100					105	
CTG	GAG	AAA	TAT	GAA	AGA	GAA	GCT	GAG	GAA	AAA	AAC	AGA	GAA	ACC	TGG	506
Leu	Glu	Lys	Tyr	Glu	Arg	Glu	Ala	Glu	Glu	Lys	Asn	Arg	Glu	Thr	Trp	
				110					115					120		
TAT	TTG	TCC	TGG	GCC	TTA	GAT	ACA	AAT	CAG	GAG	GAA	CGA	GAC	AAG	GGT	554
Tyr	Leu	Ser	Trp	Ala	Leu	Asp	Thr	Asn	Gln	Glu	Glu	Arg	Asp	Lys	Gly	
			125					130					135			
AAA	ACA	GTC	GAA	GTG	GGT	CGT	GCC	TAT	TTT	GAA	ACA	GAA	AGG	AAA	CAT	602
Lys	Thr	Val	Glu	Val	Gly	Arg	Ala	Tyr	Phe	Glu	Thr	Glu	Arg	Lys	His	
	140						145					150				
TTC	ACA	ATT	TTA	GAT	GCC	CCT	GGC	CAC	AAG	AGT	TTT	GTC	CCA	AAT	ATG	650
Phe	Thr	Ile	Leu	Asp	Ala	Pro	Gly	His	Lys	Ser	Phe	Val	Pro	Asn	Met	
	155					160					165					
ATT	GGT	GGT	GCT	TCT	CAA	GCT	GAT	TTG	GCT	GTG	CTG	GTC	ATC	TCT	GCC	698
Ile	Gly	Gly	Ala	Ser	Gln	Ala	Asp	Leu	Ala	Val	Leu	Val	Ile	Ser	Ala	
	170				175					180					185	
AGG	AAA	GGA	GAG	TTT	GAA	ACT	GGA	TTT	GAA	AAA	GGT	GGA	CAG	ACA	AGA	746
Arg	Lys	Gly	Glu	Phe	Glu	Thr	Gly	Phe	Glu	Lys	Gly	Gly	Gln	Thr	Arg	
				190					195					200		
GAA	CAT	GCG	ATG	TTT	GGC	AAA	ACG	GCA	GGA	GTA	AAA	CAT	TTA	ATA	GTG	794
Glu	His	Ala	Met	Phe	Gly	Lys	Thr	Ala	Gly	Val	Lys	His	Leu	Ile	Val	
			205					210					215			
CTT	ATT	AAT	AAG	ATG	GAT	GAT	CCC	ACA	GTA	AAT	TGG	GGC	ATC	GAG	AGA	842
Leu	Ile	Asn	Lys	Met	Asp	Asp	Pro	Thr	Val	Asn	Trp	Gly	Ile	Glu	Arg	
		220					225					230				
TAT	GAA	GAA	TGT	AAA	GAA	AAA	CTG	GTG	CCC	TTT	TTG	AAA	AAA	GTA	GGC	890
Tyr	Glu	Glu	Cys	Lys	Glu	Lys	Leu	Val	Pro	Phe	Leu	Lys	Lys	Val	Gly	
	235					240					245					
TTT	AGT	CCA	AAA	AAG	GAC	ATT	CAC	TTT	ATG	CCC	TGC	TCA	GGA	CTG	ACC	938
Phe	Ser	Pro	Lys	Lys	Asp	Ile	His	Phe	Met	Pro	Cys	Ser	Gly	Leu	Thr	
	250				255					260				265		
GGA	GCA	AAT	ATT	AAA	GAG	CAG	TCA	GAT	TTC	TGC	CCT	TGG	TAC	ACT	GGA	986

Gly	Ala	Asn	Ile	Lys	Glu	Gln	Ser	Asp	Phe	Cys	Pro	Trp	Tyr	Thr	Gly	
				270					275					280		
TTA	CCA	TTT	ATT	CCG	TAT	TTG	AAT	AAC	TTG	CCA	AAC	TTC	AAC	AGA	TCA	1034
Leu	Pro	Phe	Ile	Pro	Tyr	Leu	Asn	Asn	Leu	Pro	Asn	Phe	Asn	Arg	Ser	
			285					290					295			
ATT	GAT	GGA	CCA	ATA	AGA	CTG	CCA	ATT	GTG	GAT	AAG	TAC	AAA	GAT	ATG	1082
Ile	Asp	Gly	Pro	Ile	Arg	Leu	Pro	Ile	Val	Asp	Lys	Tyr	Lys	Asp	Met	
		300					305					310				
GGC	ACT	GTG	GTC	CTG	GGA	AAG	CTG	GAA	TCC	GGG	TCC	ATT	TTT	AAA	GGC	1130
Gly	Thr	Val	Val	Leu	Gly	Lys	Leu	Glu	Ser	Gly	Ser	Ile	Phe	Lys	Gly	
	315					320					325					
CAG	CAG	CTC	GTG	ATG	ATG	CCA	AAC	AAG	CAC	AAT	GTA	GAA	GTT	CTT	GGA	1178
Gln	Gln	Leu	Val	Met	Met	Pro	Asn	Lys	His	Asn	Val	Glu	Val	Leu	Gly	
330					335					340					345	
ATA	CTT	TCT	GAT	GAT	ACT	GAA	ACT	GAT	TTT	GTA	GCC	CCA	GGT	GAA	AAC	1226
Ile	Leu	Ser	Asp	Asp	Thr	Glu	Thr	Asp	Phe	Val	Ala	Pro	Gly	Glu	Asn	
				350					355					360		
CTC	AAA	ATC	AGA	CTG	AAG	GGA	ATT	GAA	GAA	GAA	GAG	ATT	CTT	CCA	GAA	1274
Leu	Lys	Ile	Arg	Leu	Lys	Gly	Ile	Glu	Glu	Glu	Glu	Ile	Leu	Pro	Glu	
			365				370						375			
TTC	ATA	CTT	TGT	GAT	CCT	AGT	AAC	CTC	TGC	CAT	TCT	GGA	CGC	ACG	TTT	1322
Phe	Ile	Leu	Cys	Asp	Pro	Ser	Asn	Leu	Cys	His	Ser	Gly	Arg	Thr	Phe	
		380					385					390				
GAT	GTT	CAG	ATA	GTG	ATT	ATT	GAG	CAC	AAA	TCC	ATC	ATC	TGC	CCA	GGT	1370
Asp	Val	Gln	Ile	Val	Ile	Ile	Glu	His	Lys	Ser	Ile	Ile	Cys	Pro	Gly	
	395					400					405					
TAT	AAT	GCG	GTG	CTG	CAC	ATT	CAT	ACT	TGT	ATT	GAG	GAA	GTT	GAG	ATA	1418
Tyr	Asn	Ala	Val	Leu	His	Ile	His	Thr	Cys	Ile	Glu	Glu	Val	Glu	Ile	
410					415				420					425		
ACA	GCG	TTA	ATC	TCC	TTG	GTA	GAC	AAA	AAA	TCA	GGG	GAA	AAA	AGT	AAG	1466
Thr	Ala	Leu	Ile	Ser	Leu	Val	Asp	Lys	Lys	Ser	Gly	Glu	Lys	Ser	Lys	
				430				435					440			
ACA	CGA	CCC	CGC	TTC	GTG	AAA	CAA	GAT	CAA	GTA	TGC	ATT	GCT	CGT	TTA	1514
Thr	Arg	Pro	Arg	Phe	Val	Lys	Gln	Asp	Gln	Val	Cys	Ile	Ala	Arg	Leu	
			445					450				455				
AGG	ACA	GCA	GGA	ACC	ATC	TGC	CTC	GAG	ACG	TTC	AAA	GAT	TTT	CCT	CAG	1562
Arg	Thr	Ala	Gly	Thr	Ile	Cys	Leu	Glu	Thr	Phe	Lys	Asp	Phe	Pro	Gln	

460	465	470	
ATG GGT CGT TTT ACT TTA AGA GAT GAG GGT AAG ACC ATT GCA ATT GGA			1610
Met Gly Arg Phe Thr Leu Arg Asp Glu Gly Lys Thr Ile Ala Ile Gly			
475	480	485	
AAA GTT CTG AAA TTG GTC CCA GAG AAG GAC TAAGCAATTT TCTTGATGCC			1660
Lys Val Leu Lys Leu Val Pro Glu Lys Asp			
490	495		
TCTGCAAGAT ACTGTGAGGA GAATTGACAG CAAAAGTTCA CCACCTACTC TTATTTACTG			1720
CCCATTTGATT GACTTTTTCTT CATATTTTGC AAAGAGAAAT TTCACAGCAA AAATTCATGT			1780
TTTGTCAGCT TTCTCATGTT GAGATCTGTT ATGTCACTGA TGAATTTACC CTCAAGTTTC			1840
CTTCCTCTGT ACCACTCTGC TTCCTTGGAC AATATCAGTA ATAGCTTTGT AAGTGATGTG			1900
GACGTAATTG CCTACAGTAA TAAAAAATA ATGTACTTTA ATTTTTCATT TTCTTTTAGG			1960
ATATTTAGAC CACCCTTGTT CCACGCAAAC CAGAGTGTGT CAGTGTGTTGT GTGTGTGTTA			2020
AAATGATAAC TAACATGTGA ATAAAATACT CCATTTG			2057

[Document Name] Abstract

[Abstract]

The present invention provides human genes which make it possible to detect the expression of the same in various human tissues, analyze their structures and functions, and produce the human proteins encoded by the genes by the technology of genetic engineering. Through these, it becomes possible to analyze the corresponding expression products, elucidate the pathology of diseases associated with the genes, for example hereditary diseases and cancer, and diagnose and treat such diseases.

[Means for Solution]

For example, a novel human gene comprising a nucleotide sequence coding for the amino acid sequence shown under SEQ ID NO:1.

[Drawing for Selection] None